Check for updates

ARTICLE

Artificial intelligence-assisted mapping of proliferation centers allows the distinction of accelerated phase from large cell transformation in chronic lymphocytic leukemia

Siba El Hussein [1,2,4], Pingjun Chen [3,4], L. Jeffrey Medeiros [2], John D. Hazle[3], Jia Wu[3,5✉] and Joseph D. Khoury [2,5✉]

Chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL) is characterized morphologically by numerous small lymphocytes and pale nodules composed of prolymphocytes and paraimmunoblasts known as proliferation centers (PCs). Patients with CLL can undergo transformation to a more aggressive lymphoma, most often diffuse large B-cell lymphoma (DLBCL), known as Richter transformation (RT). An accelerated phase of CLL (aCLL) also may be observed which correlates with subsequent transformation to DLBCL, and may represent an early stage of transformation. Distinguishing PCs in CLL from aCLL or RT can be diagnostically challenging, particularly in small needle biopsy specimens. Available guidelines pertaining to distinguishing CLL from its' progressive forms are limited, subject to the morphologist's experience and are often not completely helpful in the assessment of scant biopsy specimens. To objectively assess the extent of PCs in aCLL and RT, and enhance diagnostic accuracy, we sought to design an artificial intelligence (AI)-based tool to identify and delineate PCs based on feature analysis of the combined individual nuclear size and intensity, designated here as the heat value. Using the mean heat value from the generated heat value image of all cases, we were able to reliably separate CLL, aCLL and RT with sensitive diagnostic predictive values.

INTRODUCTION

Artificial intelligence (AI) algorithms have the potential to provide clinical-grade tools to assist in the diagnostic evaluation of tissue biopsy samples. Exploration of AI tools in the evaluation of lymphoid neoplasms in tissue samples has been limited. Most deep learning studies have adopted transfer learning to fine-tune an existing convolutional neural network model. In particular, the whole-slide image is divided into patches for diagnostic predictions, where the patch-wise predictions are fused to render a final diagnosis at whole-slide level. This technique has been used to distinguish benign from malignant conditions[1–5] of different types of lymphomas[4,5], or to predict the onset of large cell transformation[6].

Lymph nodes involved by chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL) are characterized by replacement of nodal architecture by a dominant infiltrate of small lymphocytes interspersed by areas termed proliferation centers (PC). The latter areas are composed of prolymphocytes and paraimmunoblasts that have increased mitotic activity. While most CLL patients have an indolent clinical course, a subset can develop more aggressive disease, either "accelerated phase of CLL/SLL" (aCLL) or Richter transformation (RT)[7]. Histologically, lymph nodes with aCLL have an increased number and size of PC, which become confluent (by definition, broader than a 20x field[7]). These changes also entail an increased Ki67 proliferation index (>40% in

PCs) and mitotic figures (>2.4 mitoses/PC)[7]. A subset of CLL patients, with or without detectable aCLL, develop disease transformation—also called RT—whose most common histologic manifestation is diffuse large B-cell lymphoma. RT is characterized by confluent growth of large cells, occasionally interspersed by remnant CLL. Patients who develop RT have a poor prognosis, with a median overall survival of <12 months despite intensive chemoimmunotherapy[8,9].

Although data are scant regarding the clinical outcomes of patients with accelerated CLL/SLL[7], these data suggest that patients with aCLL have poorer outcomes than patients with CLL/SLL. Nevertheless, the current World Health Organization classification of hematolymphoid neoplasms[10] does not provide morphologic guidelines to assess CLL cases with clinical suspicion of disease acceleration. In addition, literature pertaining to this topic is very limited, and identifying features of disease acceleration based on available limited guidelines (PCs broader than a 20x field, Ki67 proliferation index >40% in PCs and mitotic figures >2.4 mitoses/PC)[7] are subjective and depend on hematopathologist's experience, especially in scant tissue biopsy samples[7,11,12].

The application of computer-aided diagnostic algorithms based on clinically interpretable models may provide a much-needed assistance in this well defined clinical scenario. Previously, we have proposed and validated an AI model based on four morphologically meaningful cellular attributes (nuclear size, intensity, cellular

[1]Department of Pathology, University of Rochester Medical Center, Rochester, NY, USA. [2]Department of Hematopathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [3]Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [4]These authors contributed equally: Siba El Hussein, Pingjun Chen. [5]These authors jointly supervised this work: Jia Wu, Joseph D. Khoury ✉email: Jwu11@mdanderson.org; Jkhoury@mdanderson.org
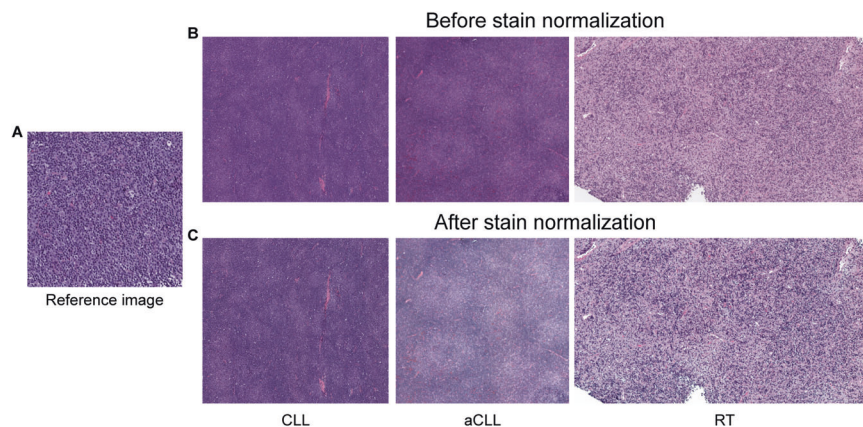
Fig. 1 **Digital slide staining normalization. A** Illustration of the reference image used for staining normalization is provided; Illustration of digital slides images from the three disease entities before **B** and **C** after staining normalization. CLL chronic lymphocytic leukemia/small lymphocytic lymphoma, aCLL accelerated chronic lymphocytic leukemia/small lymphocytic lymphoma, RT Richter transformation, diffuse large B-cell lymphoma variant.
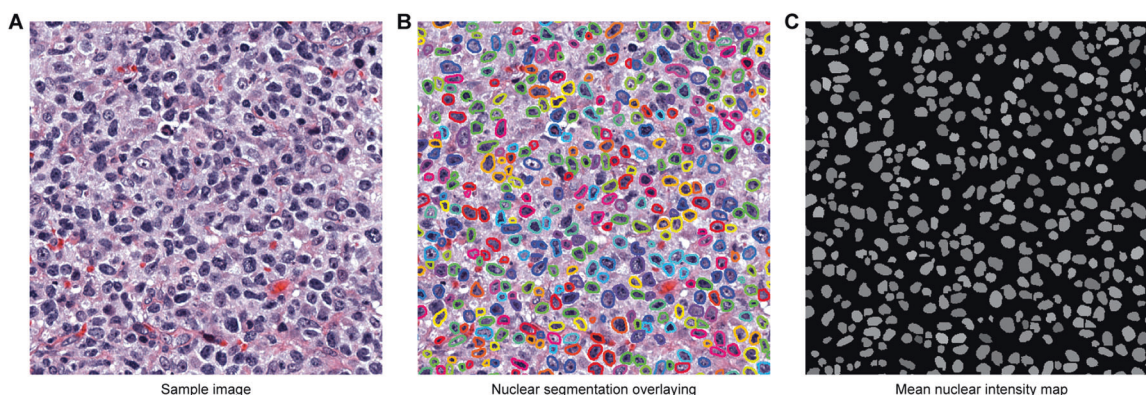

Fig. 2 **Nuclear segementation. A** Example image from a Richter transformation, diffuse large B-cell lymphoma (RT) case, **B** with application of nuclear segmentation, followed by **C** generation of nuclear intensity map.

density and cell to nearest-neighbor distance) to help distinguish CLL, aCLL and RT, showing a satisfactory predictive accuracy[13].

In this study, we sought to design an AI-based tool that can provide an objective assessment to understanding low-power magnification architectural changes, and for enhancing the delineation of PCs in CLL, and in its accelerated and transformed phases. To this aim, we have performed a combined "nuclear size and intensity analysis" that we termed "heat value". Using the mean heat value from the generated heat value image of all cases, we were able to reliably separate the three phases in question with sensitive diagnostic predictive values.

MATERIAL AND METHODS
Data collection
The study was approved by The University of Texas MD Anderson Cancer Center (MDACC) Institutional Review Board and conducted in accord with the Declaration of Helsinki. We retrospectively searched patients with hematolymphoid diseases clinically evaluated at MDACC between February 1, 2009, and July 31, 2021. We randomly selected 10 CLL, 12 aCLL, and 8 RT digitized hematoxylin and eosin stained slides of excisional biopsy specimens of lymph nodes to study the mapping of PC. Slide scanning was conducted using Aperio AT2 scanners at an optical resolution of ×20 (0.50 µm/pixel). All selected slides came from different patients, and in total we manually annotated 25, 28, and 21 regions of interest (ROI) encompassing small round PCs and confluent/expanded PCs from CLL, aCLL, and RT, respectively. ROI selection was random in the three disease categories and did not target any specific areas to decrease selection bias, however during ROI annotation, we avoided areas with tissue folding, red blood cell extravasation and accumulation, and in RT cases

areas of necrosis, as these morphologic features could affect the performance of our algorithm targeting our cells of interest. To ensure the heatmap generated from mapping of PC had sufficient information, both the length and width of the annotated ROI were required to be larger than 2000 pixels. Stain normalization was performed on all ROIs prior to further processing[14] (Fig. 1).

Cell segmentation and refinement
This study aimed to objectively automate mapping of PCs, which is visually delineated by pathologists in clinical practice during evaluation of glass slides. Our proposed mapping model is based on visual properties of individual nuclear size and intensity, thus nuclear segmentation is a prerequisite. We employed Hover-Net for the nuclear segmentation task, given that Hover-Net is a state-of-the-art nuclear segmentation algorithm pre-trained on the MoNuSeg dataset, thus avoiding time-consuming nuclear annotation and model tuning procedures[15,16]. We performed a quantitative evaluation of 15 manually annotated patches of 256×256 pixels and achieved an overall mean Dice score of 0.825. The Hover-Net had Dice scores of 0.826 and 0.853 on datasets Kumar and CoNSep, respectively. Nuclear segmentation of our dataset was consistent with Hover-Net's reported performance. A strong agreement between the manual and Hover-Net segmentations was indicated with a Dice score over 0.80, thus laying the foundation for the proposed nuclear-based feature engineering. The Hover-Net results were also visually checked by a hematopathologist (S. E.H.) to assure nuclear segmentation quality (Fig. 2A, B).

Nevertheless, after Hover-Net segmentation, inevitably we encountered few overlapping nuclei that were inaccurately segmented. To address this issue, we deployed the solidity feature, defined as the ratio of segmented nuclear contour area to its convex hull area, to filter out overlapping nuclei. Segmented nuclei with a solidity value smaller than 0.84 were removed from further analysis. In addition, we set the minimum and maximum pixel number of the
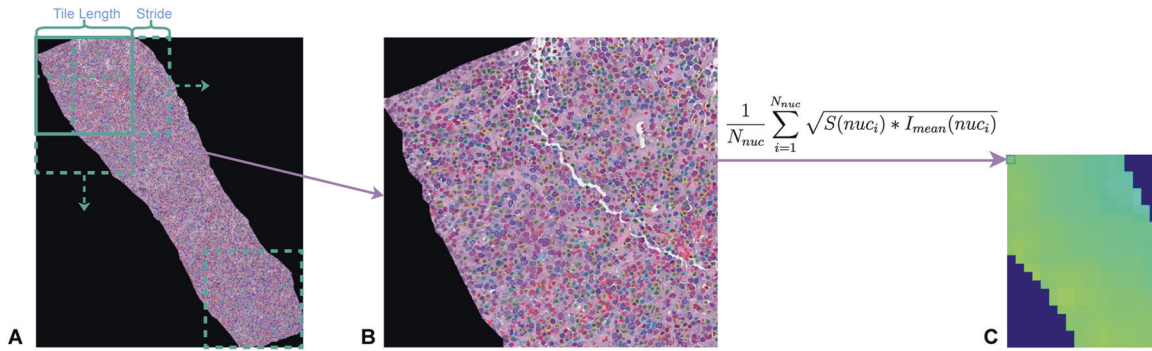
Fig. 3 Generation of heat values by integrating nuclear size and intensity analysis. A Regions of interest (ROIs) are split into multiple tiles by setting two parameters, the tile strength and stride; **B** this is followed by feature analysis of the combined size/intensity of segmented nulcei inside each tile; **C** the estimated heat value per tile is generated by integrating nuclear size and mean intensity using the following formula: $\frac{1}{N_{nuc}} \sum_{i=1}^{N_{nuc}} \sqrt{S(\text{nuc}_i) \times I_{mean}(\text{nuc}_i)}$. nuc nuclear, S size, I intensity.

nuclei to be 32 and 432, which corresponds to 8 and 108 μm², respectively. We also discarded nuclei with a pixel number outside the set range.

Automated mapping of proliferation centers

We first split each ROI into multiple tiles by setting two parameters, the tile length and stride (Fig. 3A). These two parameters are set to be the same on both horizontal and vertical directions and across all ROI. In our experiments, we set the tile length and stride as 1000 and 100 pixels, respectively. As the value of tile length is larger than the stride, some neighboring tiles overlapped, thus creating a sufficient number of tiles to map PCs.

We then conducted feature analysis of the combined size/intensity properties (Fig. 2C) of nuclei inside each tile (Fig. 3B), to generate and recreate a novel representation of PCs. As nuclear size varied from 8 to 108 square micrometers, and nuclear mean intensity varied from 0 to 255, we normalized the values of nuclear size and mean intensity to 0.0 and 1.0, by subtracting the minimum value and dividing it by the value range length, and called them $S(\text{nuc}_i)$ and $I_{mean}(\text{nuc}_i)$, respectively. We then estimated the heat value of each tile by integrating nuclear size and mean intensity using the following formula:

$$\frac{1}{N_{nuc}} \sum_{i=1}^{N_{nuc}} \sqrt{S(\text{nuc}_i) \times I_{mean}(\text{nuc}_i)}.$$

With the proposed heat value estimation formula, we calculated the heat value for each tile in each ROI. We then generated a heat value image for each ROI to map its PCs and repeated this process for ROI of all CLL, aCLL, and RT cases (Fig. 4A–D).

Based on our proposed heat value formula, the heat values in the heat value image were less than 0.5 in analyzed ROIs, thus the contrast of the generated heat value image is very limited. Instead of directly converting the heat value image to the heatmap, we first accentuated heat values by multiplying by a factor of 2.0 to increase the contrast, followed by heatmap conversion based on the scaled heat value image for better visualization. By multiplying the heat values by a factor of 2.0, the resulted heat values were still in the range of 0.0 to 1.0, without value saturation. We applied Otsu's method to identify the optimal threshold for each heat value image and regarded the segmented foreground as PC areas.

For an objective quantification of the heat value image, we went a step further and generated a heat value histogram for each heat value image (Fig. 4E). Based on the obtained histograms, we employed the F-score (a measure of a test's accuracy using the following formula: $F\text{-score} = 2.0 \times (\text{precision} \times \text{recall})/(\text{precision} + \text{recall})$) to identify two heat value thresholds and achieve an optimal separation performance among the three entities (Supplementary Fig. 5). The two-sided Welch's t test was used to quantify the difference between diseases. Furthermore, we employed the mean value from the ROI heat value histogram to evaluate the diagnostic performance (Supplementary Fig. 6).

RESULTS

Heatmaps were generated based on heat values per tile inside each ROI of the three disease phases (CLL, aCLL, and RT), as illustrated in

Fig. 4A–D. The intensity of the heat value image (Fig. 4B) was accentuated to sharpen color separation, then converted to a heatmap (Fig. 4C), to recreate the PCs (Fig. 4D). Areas with high heat values (in the yellow spectrum) correspond to tiles harboring cells with increased nuclear size and mean intensity (PCs in CLL cases and expanded/confluent PCs in aCLL and RT cases) (Fig. 4D). In contrast, areas with low heat values (in the blue spectrum) correspond to tiles with decreased nuclear size and mean intensity, representing small neoplastic lymphocytes surrounding PCs (Fig. 4D). This recreation of PCs based on objective measures of nuclear attributes (size and intensity) provides on its own a visual aid to assess the extent of large cells (with large nuclei) depicted in yellow in relation to small neoplastic lymphocytes in blue, in the three disease phases: Yellow foci confined in small PC, and occupying a subset of the ROI, with predominantly blue areas composed of small-size neoplastic lymphocytes with decreased intensity in CLL; Expanding yellow foci creating confluent PCs, with decreasing background blue areas in aCLL; Predominantly yellow ROI with sheets of large cells, resulting from fusing of PCs, and virtually absent blue areas in RT (Fig. 4D).

Based on obtained histograms from heat value images (Fig. 4E), we isolated two optimal heat value thresholds based on the F-score to achieve the best separation among the three disease phases: 0.228, below which the case was most likely to be CLL (the top histogram in Fig. 4E represents a CLL ROI with heat values ranging between 0.16 and 0.19, and peaking at 0.18 approximately); and 0.288, above which the case was most likely to be RT (the bottom histogram in Fig. 4E represents an RT ROI with heat values ranging between 0.20 and 0.30, and peaking at 0.27 approximately); Cases with heat values ranging between 0.228 and 0.288 were most likely aCLL (the middle histogram in Fig. 4E represents an aCLL ROI with heat values ranging between 0.28 and 0.35, and peaking at 0.29 approximately). We then plotted the mean heat value frequencies of all ROI from the three phases: There was a significant difference in the ranges of mean heat value frequencies for CLL, aCLL, and RT, which were 0.168 to 0.233, 0.212 to 0.307, and 0.261 to 0.353, respectively (Supplementary Fig. 5).

Besides, we performed a diagnostic study to evaluate the mean heat value's prediction ability via repeated splitting analysis, where we randomly split the whole dataset into training and testing cohorts 100 times, stratified at patient level with a ratio of 1:1. The splitting was performed patient-wise, insead of ROI-wise, to avoid selecting ROIs belonging to the same patients in both the training and testing sets. The accuracy and area under the curve diagnostic predictive values using data from nuclear size alone were 0.658 (±0.115) and 0.771 (±0.096), respectively; and using mean nuclear intensity, 0.679 (±0.094) and 0.841 (±0.052), respectively; with a noticeable increase using heat value frequencies (integrating the nuclear size and mean nuclear intensity) reaching 0.813 (±0.0630) and 0.885 (±0.109), respectively (Supplementary Fig. 6).
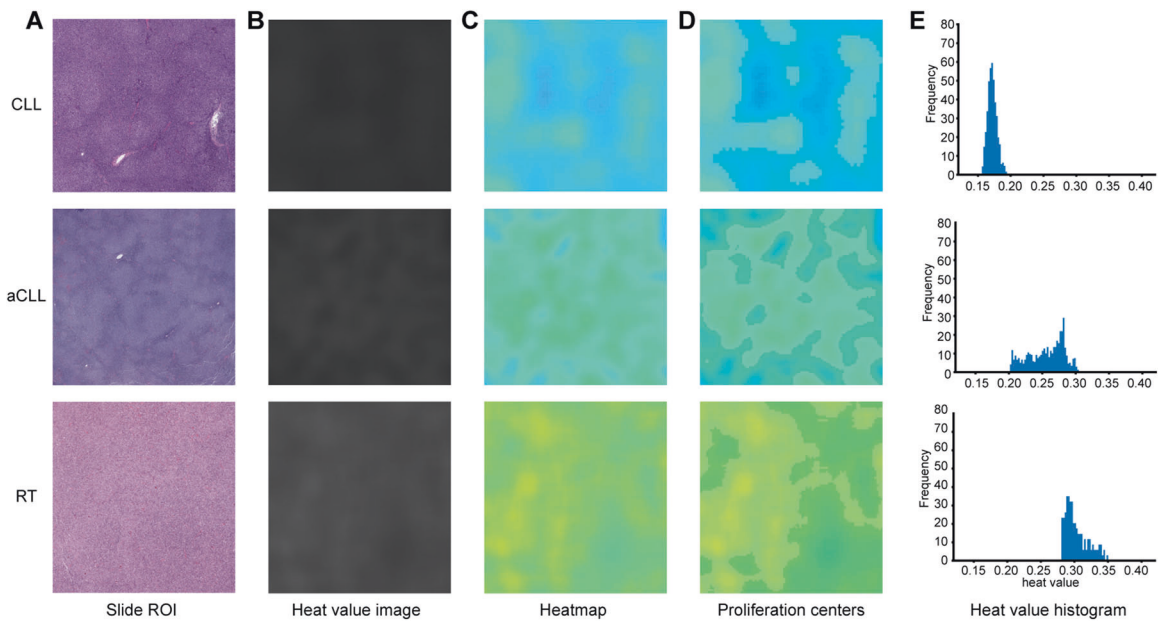
Fig. 4 Heatmap generation based on heat values per tile for each of the three disease entities. A Based on image obtained from selected regions of interest (ROIs), **B** heat value images are generated, **C** followed by heat map generation and **D** accentuation to re-create proliferation centers. Areas with high heat value frequencies (yellow spectrum) correspond to tiles harboring cells with increased nuclear size and mean intensity (PCs in CLL cases and expanded/confluent PCs in aCLL and RT cases). In contrast, areas with low heat value frequencies (blue spectrum) correspond to tiles with decreased nuclear size and mean intensity (small neoplastic lymphocytes surrounding PCs); **E** a heat value frequency histogram per tile per ROI is then created for each case: the two optimal thresholds isolated to obtain the highest separation value among the three disease entities were: 0.228, below which the case was most likely to be CLL, and 0.288 above which the case was most likely to be RT. Cases with heat value frequencies ranging between 0.228 and 0.288 were most likely aCLL cases.

DISCUSSION

Clinically, accelerated phase disease behaves similarly to typical CLL in terms of B-symptoms, disease bulk, functional status and clinical stage, these patients often have higher serum LDH levels and their lymphoma is ZAP70 positive. Some data suggest that the prognosis of aCLL patients is poorer than that of CLL patients. In contrast, patients with unequivocal disease transformation, or RT, are notoriously more symptomatic, have lower performance status, higher serum LDH levels, and higher uptake on PET-CT scan[17]. Although in some practices, aCLL cases are still treated like CLL (combination therapy of ibrutinib and venetoclax)[17], switching into a more intensive treatment regimen in patients with aCLL is deployed in some settings with better clinical response, especially in CLL patients who become refractory to treatment. Thus the importance of distinguishing classic CLL from its accelerated phase morphologically, corroborating with clinical suspicion of disease progression and the need to upgrade treatment. However data from prospective clinical trials is needed to better assess the long term benefits/outcomes of aggressive therapeutic strategies in aCLL.

Hematopathologists rely on low magnification microscopic examination to characterize the shape of PCs in patients with history of CLL. Small round and distinct PCs are indicative of an underlying CLL, whereas confluent PCs occupying larger areas are more indicative of underlying disease acceleration (aCLL). Lastly, expanded sheets of large cells, beyond a recognizable PC morphology, is diagnostic of RT. Analysis of PC expansion/ formation of sheets of large cells is conducted based on the assessment of H&E glass slide at low magnification, coupled with Ki67 stain that may highlight the extent of large cell (~mitotically active cell) expansion. Gine et al. defined expanded PC in aCLL as broader than a 20x field[7]. However, in our experience, this assessment is morphologist-dependent and varies greatly depending on the exposure of hematopathologists to these particular cases[9].

We previously published an AI-based "disease diagnosis model" in which we isolated "nuclear/cellular morphologic features" that we implemented as biomarkers to enhance diagnostic accuracy in CLL, aCLL and RT[13]. In the present study, we sought to design an "architecture-based" tool to enhance the delineation of PCs, by implementing a novel technique that integrates nuclear size and intensity. By applying this tool, large nuclei (~large cells) with high intensity, and small nuclei (~small cells) with low intensity occupy the yellow and blue spectra, respectively (Fig. 4A–D). Using this method, we were able to enhance the visualization of large cell overall architectural distribution and extent in studied ROI: confined yellow foci in PCs in CLL, confluent yellow foci representing fused PCs in aCLL, or yellow sheets replacing the vast majority of the ROI in RT (Fig. 4D). As we were able to reproduce these results across ROI from the three disease phases, we propose that this tool, with further fine-tuning, could be implemented in the future to further assist in the visual assessment of challenging cases with features of disease acceleration/transformation, especially in limited core-needle biopsy specimens.

In addition to visually mapping the extent of large cells, we plotted the heat values of all tiles to their frequencies per ROI (Fig. 4E). This technique is likened to a cell size and intensity "sorter". By repeating this process in all ROIs, we isolated two heat value coefficients: 0.228, below which a case is most likely to be classified as CLL; and 0.288, above which a case is most likely to be classified as RT. Cases lying in between these two coefficients are most likely to be aCLL. We also noticed that heat value frequencies in CLL ROI had a single Gaussian-like prominent peak, as illustrated in the example we provide in the top histogram in Fig. 4E. Heat value frequencies that demonstrated a smearing pattern, with no definitive dominating peak in aCLL cases, are illustrated in the example we provide in the middle histogram in Fig. 4E. Finally, RT cases demonstrated a right-shifted distribution of heat value distribution frequency, with some tiles occupying increasing heat values, as illustrated in the example we provide in the bottom histogram in Fig. 4E.

To test the generalizability of the findings described above, we plotted the mean heat value in all ROIs (y axis) across the three

disease phases (x axis). The range of mean heat value frequencies demonstrated a statistically significant separation among the three phases in most cases, with one aCLL case overlapping with RT range, and a few CLL cases overlapping with aCLL range (Supplementary Fig. 5). Overlapping ranges among entities on rare occasions could be attributed to ROI sampling of active PCs belonging to the far end side of disease spectrum, as PCs in CLL and aCLL are dynamic environments and could be captured at any point during their growth, including phases in which a minor percentage of them have crossed over to a more progressive state (CLL into aCLL or aCLL into RT). In the future, the robustness of our model will need to be tested in a multicenter environment with different pathologists.

Our data suggest that this model, based on objective architectural analysis of PCs, is able to achieve a high diagnostic accuracy. Although our design was performed on excisional biopsy specimens, which are inherently more informative morphologically, the end goal of this model is to deploy it in limited biopsy specimens. In fact, core-needle biopsy is nowadays a more common method of tissue sampling in the setting of clinical suspicion of underlying disease progression/transformation, as these specimens can be obtained more rapidly and are minimally invasive in comparison to excisional biopsies. However, core-needle biopsy specimens provide an incomplete picture of the underlying nodal architecture, a keystone in the assessment of accelerated disease, and delivering an accurate and confident diagnosis in this challenging scenario may be achieved by the assistance of objective tools. We suggest that our model, with further refinement and sophistication, can be ultimately deployed to this aim.

In summary, our study provides an architecture-based tool to objectively assess the extent of PCs in CLL cases with clinical suspicion of disease progression, based on the integrative analysis of cell nuclear size and mean nuclear intensity and automation of PC mapping. Using the mean heat value of all cases, we were able to reliably separate the three disease phases in question with sensitive diagnostic predictive values. We suggest that an ROI mean heat value less than 0.228 is predictive of CLL, and a value more than 0.288 is predictive of RT. aCLL cases demonstrate a mean heat value ranging from 0.228 to 0.288. These thresholds need to be independently verified using external image sets to ensure generalizability. Nevertheless, this work highlights the value of using AI-based tools in identifying clinically meaningful cellular and architectural features, to enhance disease diagnosis in challenging clinical scenarios. Our model, although trained and tested on excisional biopsy specimens, could be potentially very useful in the assessment of limited core-needle biopsy specimens, where typically only a small percentage of PCs is available for morphologic evaluation of architecture and extent of growth.

DATA AVAILABILITY
All the original data of this study will be available upon reasonable request to the corresponding authors, including, but not limited to, a request to reproduce results in this manuscript.

REFERENCES

1. Syrykh, C. et al. Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning. *NPJ Digit. Med.* **3**, 63 (2020).
2. Li, D. et al. A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals. *Nat. Commun.* **11**, 6004 (2020).
3. Miyoshi, H. et al. Deep learning shows the capability of high-level computer-aided diagnosis in malignant lymphoma. *Lab. Invest.* **100**, 1300–1310 (2020).
4. Achi, H. E. et al. Automated diagnosis of lymphoma with digital pathology images using deep learning. *Ann. Clin. Lab. Sci.* **49**, 153–160 (2019).
5. Mohlman, J. S. et al. Improving augmented human intelligence to distinguish Burkitt lymphoma from diffuse large B-cell lymphoma cases. *Am. J. Clin. Pathol.* **153**, 743–759 (2020).
6. Irshaid, L. et al. Histopathologic and machine deep learning criteria to predict lymphoma transformation in bone marrow biopsies. *Arch. Pathol. Lab. Med.* https://doi.org/10.5858/arpa.2020-0510-OA (2021).
7. Gine, E. et al. Expanded and highly active proliferation centers identify a histological subtype of chronic lymphocytic leukemia ("accelerated" chronic lymphocytic leukemia) with aggressive clinical behavior. *Haematologica* **95**, 1526–1533 (2010).
8. Parikh, S. A., Kay, N. E. & Shanafelt, T. D. How we treat Richter syndrome. *Blood* **123**, 1647–1657 (2014).
9. Rossi, D., Spina, V. & Gaidano, G. Biology and treatment of Richter syndrome. *Blood* **131**, 2761–2772 (2018).
10. Swerdlow, S. H. et al. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Revised 4th edn, Vol. 2 246–247 (IARC, 2017).
11. Falchi, L. et al. Correlation between FDG/PET, histology, characteristics, and survival in 332 patients with chronic lymphoid leukemia. *Blood* **123**, 2783–2790 (2014).
12. Ciccone, M. et al. Proliferation centers in chronic lymphocytic leukemia: correlation with cytogenetic and clinicobiological features in consecutive patients analyzed on tissue microarrays. *Leukemia* **26**, 499–508 (2012).
13. El Hussein, S. et al. Artificial intelligence strategy integrating morphologic and architectural biomarkers provides robust diagnostic accuracy for disease progression in chronic lymphocytic leukemia. *J. Pathol.* https://doi.org/10.1002/path.5795 (2021).
14. Reinhard, E., Ashikhmin, M., Gooch, B. & Shirley, P. Color transfer between images. *IEEE Computer Graph Appl.* **21**, 34–41 (2001).
15. Graham, S. et al. Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* **58**, 101563 (2019).
16. Gamper, J. PanNuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. *European Congress on Digital Pathology* (Springer International Publishing, 2019).
17. Loghavi, S., Mirza, K., Wang, A. & Smith, S. M. Chronic lymphocytic leukemia/small lymphocytic lymphoma—accelerated phase. *The Pathologist.* https://thepathologist.com/subspecialties/chronic-lymphocytic-leukemia/small-lymphocytic-lymphoma. (2020).

AUTHOR CONTRIBUTIONS
S.E.H., P.C., J.W., and J.D.K. conceptualized the idea. S.E.H. and P.C. wrote the initial version of the manuscript. J.W. and J.D.K. supervised experimentation and re-iterations, and created the final version of the manuscript. L.J.M. and J.D.H. provided critical evaluation of the final version of the manuscript.

FUNDING

ETHICS APPROVAL AND CONSENT TO PARTICIPATE
All coauthors consented to participate.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41379-022-01015-9.

Correspondence and requests for materials should be addressed to Jia Wu or Joseph D. Khoury.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.