ORIGINAL ARTICLE

ISAC CYTOMETRY
Journal of Quantitative Cell Science PART A

Quantitative image analysis pipeline for detecting circulating hybrid cells in immunofluorescence images with human-level accuracy

Robert T. Heussner[1] | Riley M. Whalen[2] | Ashley Anderson[2] |
Heather Theison[2] | Joseph Baik[1] | Summer Gibbs[2,3] | Melissa H. Wong[2,3] |
Young Hwan Chang[1,3]

[1]Department of Biomedical Engineering, Oregon Health & Science University, Portland, Oregon, USA

[2]Department of Cell, Developmental and Cancer Biology, Oregon Health & Science University, Portland, Oregon, USA

[3]Knight Cancer Institute, Oregon Health & Science University, Portland, Oregon, USA

Correspondence
Melissa H. Wong and Young Hwan Chang, Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA
Email: wongme@ohsu.edu and chanyo@ohsu.edu

Abstract

Circulating hybrid cells (CHCs) are a newly discovered, tumor-derived cell population found in the peripheral blood of cancer patients and are thought to contribute to tumor metastasis. However, identifying CHCs by immunofluorescence (IF) imaging of patient peripheral blood mononuclear cells (PBMCs) is a time-consuming and subjective process that currently relies on manual annotation by laboratory technicians. Additionally, while IF is relatively easy to apply to tissue sections, its application to PBMC smears presents challenges due to the presence of biological and technical artifacts. To address these challenges, we present a robust image analysis pipeline to automate the detection and analysis of CHCs in IF images. The pipeline incorporates quality control to optimize specimen preparation protocols and remove unwanted artifacts, leverages a β-variational autoencoder (VAE) to learn meaningful latent representations of single-cell images, and employs a support vector machine (SVM) classifier to achieve human-level CHC detection. We created a rigorously labeled IF CHC data set including nine patients and two disease sites with the assistance of 10 annotators to evaluate the pipeline. We examined annotator variation and bias in CHC detection and provided guidelines to optimize the accuracy of CHC annotation. We found that all annotators agreed on CHC identification for only 65% of the cells in the data set and had a tendency to underestimate CHC counts for regions of interest (ROIs) containing relatively large amounts of cells (>50,000) when using the conventional enumeration method. On the other hand, our proposed approach is unbiased to ROI size. The SVM classifier trained on the β-VAE embeddings achieved an F1 score of 0.80, matching the average performance of human annotators. Our pipeline enables researchers to explore the role of CHCs in cancer progression and assess their potential as a clinical biomarker for metastasis. Further, we demonstrate that the pipeline can identify discrete cellular phenotypes among PBMCs, highlighting its utility beyond CHCs.

KEYWORDS

cancer metastasis, cell phenotyping software, circulating hybrid cells, machine learning

1 | INTRODUCTION

Cancer is the second leading cause of death globally, and despite notable advancements in treatment, significant challenges persist [1]. Timely information regarding disease burden is crucial for early detection, monitoring treatment response, and assessing the risk of metastasis or recurrence. However, current state-of-the-art assays and approaches for procuring such information pose difficulties. Biopsies and molecular analysis of tumor samples offer valuable insights but are invasive, limited in scope, and not feasible after tumor resection [2]. Alternatively, blood-based biomarkers hold promise, as they can be monitored from a peripheral blood sample, which is far less invasive and facilitates more frequent monitoring during treatment. Current blood-based cancer biomarkers include non-cell-based biomarkers (i.e., cell-free nucleotides, proteins, exosomes) and cell-based biomarkers [3]. Non-cell-based cancer biomarkers, such as carcinoembryonic antigen (CEA) in colorectal cancer, are used for monitoring disease progression [4], but are not always accurate nor detected in every patient. Further, other biomarkers such as cell-free DNA (cfDNA) offer one-dimensional information about the disease. Only small amounts of genomic information can be gained from cfDNA without additional tumor-associated protein information, and vice versa for protein-based biomarkers such as CEA [3]. Cell-based biomarkers are directly disseminated from the tumor and allow for a wider range of information to be assessed (i.e., RNA, genomic, and proteomic data) [3]. Circulating tumor cells (CTCs) are the most well-researched cell-based biomarker for cancer [5–9]. However, they are rare in peripheral blood (less than 5 CTCs per 7.5 mL of blood in high-burden settings) [6], which makes them difficult to isolate and analyze, and their numbers do not always correlate with patient outcomes [5–9].

Circulating hybrid cells (CHCs) are a newly discovered population of disseminated tumor cells that co-express immune and neoplastic proteins and show promise as a cell-based biomarker [10–12]. CHCs can result from a fusion event between an immune cell and a tumor cell, granting it novel traits, such as enhanced migratory properties, and increased metastatic potential [10]. CHCs are identified in humans and murine tumor models in numbers that exceed that of CTCs in peripheral blood [12–20] across various types of cancers. Additionally, the number of CHCs in a cancer patient's peripheral blood correlates with disease stage and outcome, solidifying their potential as a useful cancer biomarker [10–12, 21]. To optimize the utility of this discovery, the development of an accessible platform capable of identifying and analyzing CHCs is required. Such a platform and subsequent analytics will provide the foundation to move CHC analyses from research-based discovery to clinically approved assays, and would improve cancer management by supplying valuable and actionable information for personalized and timely treatment decisions [5].

Several FDA-approved commercial platforms are available to detect CTCs, such as CellSearch® [22], Vortex [23], and DEPArray [24]. CellSearch® utilizes immunomagnetic isolation, Vortex employs size-based microfluidic separation, and DEPArray utilizes

dielectrophoresis to trap cells. However, these methods focus on CTCs, which are defined as pan-leukocyte antigen negative and pan-cytokeratin positive (CD45$^-$/panCK$^+$), and exclude cells expressing immune surface antigens. This restricts their ability to identify CHCs [12, 25], which are defined as cells co-expressing CD45$^+$/panCK$^+$. Immunofluorescence (IF) imaging is predominantly used for CHC detection in laboratory settings. This approach requires labeling peripheral blood mononuclear cells (PBMCs) with panels of antibodies against cancer- and immune-associated epitopes (**Methods**). IF imaging provides a cost-effective approach for detecting CHCs, and advanced multiplex imaging techniques like cyclic IF (CyCIF) [26] allow for the measurement of more than 20 biomarkers, compared with 3–5 biomarkers in traditional IF, enabling detailed phenotyping of cells [12, 27].

Despite being an exciting, emerging biomarker, the CHC lacks a robust detection platform, particularly in the realm of image-based assays. The exploration of machine learning (ML)-based approaches for CHC detection is limited, and the development of such an ML approach for IF images presents several challenges. First, ML-based approaches struggle with an imbalanced cellular data set [28], as CHCs are relatively rare and typically constitute less than 1% of PBMCs, as shown in Figure 1A. This setback requires data curation to overcome and makes generating labeled data and discovering CHC sub-populations difficult. Second, IF images contain autofluorescence, variable intensity levels across samples or batches, and artifacts that heavily impact downstream analyses, necessitating robust quality control and normalization steps during image processing [29, 30]. Few studies address batch variation in serum samples, and a standard IF imaging protocol for serum samples is yet to be established. The lower dynamic range of biomarkers compared to tissue complicates computational thresholding methods [31].

The traditional approach to CHC detection in IF images involves manual gating of protein expression with microscopy software such as ZEISS ZEN or ImageJ [32]. However, this approach is not only time-consuming, but the extent of inter-observer and intra-observer bias in CHC detection remains unknown (Figure 1B). One CTC-based study with 11 participants reported an initial inter-observer agreement of 85% and improved it to 93% with further training [33]. Although several software solutions strive for fully automated CTC detection in IF images, most exclude cells that express CD45, making them unsuitable for CHCs. Many of them require specialized tools like functionalized and structured medical wire [34] or microfluidic devices [35] for CTC collection as a means of data curation, making them unsuitable for users lacking access to such technologies. Given the novelty of the CHC population, approaches that rely on a limited number of annotators to produce ground truth labels are undesirable [36]. Furthermore, many ML approaches directly apply classifiers to single-cell images, limiting their ability to discover new CHC phenotypes [37, 38]. Support vector machine (SVM) classifier ALICE can enumerate 20 CTC phenotypes, including CHCs [39]. However, ALICE's threshold-based segmentation is not robust against known challenges in PBMC IF images, such as artifacts and cell clumps. It relies on first capturing potential CTCs with TU-chip™, a microfluidic
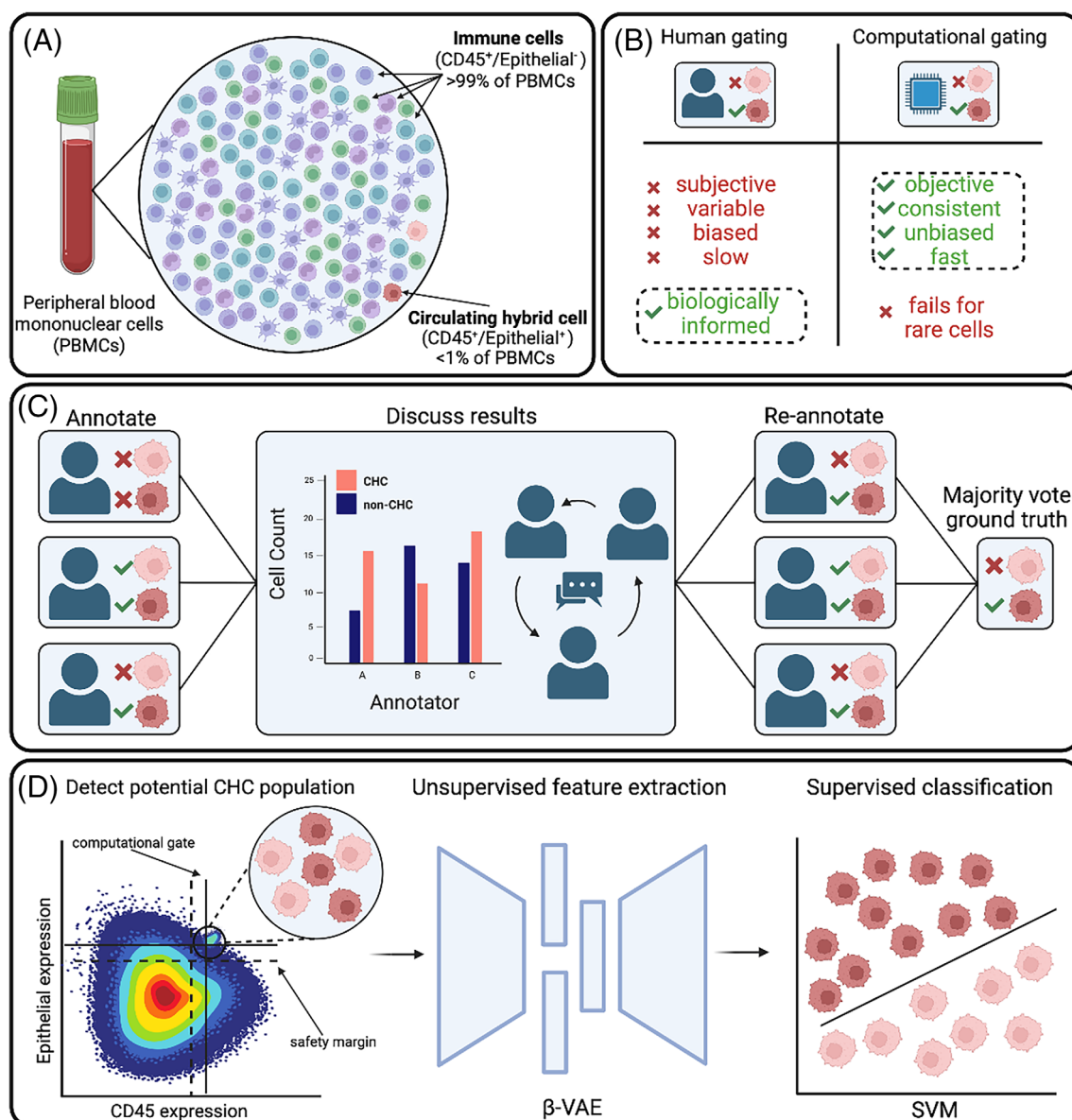
FIGURE 1 Pipeline motivation and overview. (A) CHCs are a rare cell population in peripheral blood. (B) The benefits and drawbacks of human and CG methods for inferring cell types in IF images. (C) Our rigorous, iterative human annotation study to determine GT labels. (D) Our image analysis pipeline uses CG to filter potential CHCs to a β-variational autoencoder (β-VAE) representation learner, which extracts latent features from the single-cell images to be classified with a support vector machine (SVM). [Color figure can be viewed at wileyonlinelibrary.com]

device for sorting cells by size, and has been tested for CHC detection on only pancreatic cancer data sets. Given that the size of CHCs is similar to leukocytes [10], it is most likely that the TU-chip™ is missing the majority of CHCs. Further, ALICE does not address human inter-observer and intra-observer bias in ground truth (GT) generation.

Herein, we introduce an automated CHC detection pipeline that achieves human-level performance. Our pipeline combines statistics-driven computational gating (CG) and representation learning to effectively identify and phenotype CHCs. We conducted a thorough human annotation study to establish accurate GT CHC labels. To manage the large initial data set, encompassing one million PBMCs, spanning nine patient samples and two disease sites (Table S1), we first employed CG to narrow the analytical data set to 2834 PBMCs for

human annotators to review (Figure 1C). We provided the single-cell data in a format that aligns with the annotator's familiar view of stained cells, including negative controls. We examined annotator variation and bias through a systematic evaluation of seven annotator's performances. The initial results were shared with the annotators, allowing them to learn from any gating mistakes and phenotype-calling biases. Annotators reviewed the data and developed guidelines for CHC identification. To establish the robust GT data set, we then carried out a second round of individual annotations with the original seven annotators and three additional annotators, all of whom attended the review session (Figure 1C). We used a β-variational auto-encoder (VAE) to extract latent features from the single-cell images and trained an SVM classifier to detect the CHCs (Figure 1D). Finally,

we compared the PBMC classification performance of CG alone, the SVM, and the human annotators.

2 | RESULTS

2.1 | Image processing and computational gating strategy

The primary obstacle to developing a robust ML model for CHC detection lies in obtaining a well-annotated data set. Unfortunately, there is currently no publicly available PBMC data set with adequate CHC annotations. To overcome this issue, we created our own labeled data set, enlisting 10 individuals experienced in annotating CHCs in IF images. Our collected PBMC data set consists of samples from nine patients, featuring 15 regions of interest (ROIs) and a total of 1,083,196 PBMCs. Annotating such a vast number of cells with the manual efforts of 10 people would have been impractical. To overcome this challenge, we implemented deep-learning based cell and nuclei segmentation, Mesmer [40], feature extraction, and a CG strategy (Figure 1D) to filter the data set by the expression pattern of an epithelial-directed antibody cocktail containing antibodies against E-cadherin (ECAD), epithelial cellular adhesion molecule (EpCAM), and panCK, as well as CD45, and cellular morphology, allowing annotators to focus solely on potential CHCs. By applying this filtering approach, we reduced the data set to 2834 PBMCs, accounting for only 0.26% of the original cell population. Each annotator assessed the data on a per-ROI basis (**Methods**). Implementing this filtering strategy led to a remarkable reduction in annotation time, saving hours of manual effort and enabling annotators to concentrate more on evaluating the protein expression patterns of the potential CHCs.

2.2 | Annotation study

Creating a labeled data set allowed us to assess inter-annotator variation and annotator bias. We conducted two rounds of annotations to combat the effect of annotator disagreement. We chose majority vote aggregation (6 out of 10 annotators agree) from round two to define the GT labels (Figure 1C).

Annotator F1 scores showed a significant improvement between rounds one and two, as seen in Figure 2A: For each ROI, median F1 scores increased and the interquartile range decreased. Figure S1 shows annotators tended to be more conservative in identifying CHCs in round two. Figures S2 and S3 further elucidate the inter-annotator variability in changes in F1 score and CHC counts across annotation rounds. Our findings demonstrate that after training (**Methods**), annotators achieved a higher level of consensus on cell type assignments, increasing from 40.4% to 65.7% of the data set, as shown in Figure 2B. This improvement highlights the importance of training in enhancing the consistency and accuracy of annotations.

Next, we assessed annotator bias introduced from patient-level and ROI-level variations (Figure 2C). We observed that, on average, the annotators' cell type assignments changed for 20% of the cells duplicated across different patients (**Methods**). This outcome was expected and likely influenced by the batch effect present among patients with varying disease states and data acquisition factors such as processing time, staining, and imaging. Surprisingly, we also discovered that, on average, the annotators' cell type assignments changed for 11% of the cells duplicated across different ROIs within the same patient slide. This level of ROI bias presents significant challenges to the conventional control well-based gating approach, which involves defining IF contrast limits for multiple ROIs of the same patient slide based on a single, unstained ROI to estimate autofluorescence levels.

While single-cell accuracy is important for biological investigations and cell phenotyping, the clinical use of CHCs to infer disease state and trajectory primarily relies on CHC enumeration [10, 11, 21, 41]. Consequently, we compared CG against the GT counts of both annotation rounds to assess its performance in capturing the overall trend of CHC presence within the ROIs. We use the majority vote (four out of seven annotators agree) from round one as the GT labels for comparison purposes. The results give an R^2 of 0.68 and a slope of 0.87 for round two, as shown in Figure 2D, indicating a substantial alignment between CG and GT counts. However, CG had a superior regression to the round one GT with an R^2 of 0.77, highlighting the necessity for more advanced classification methods, especially in light of the observed elevated F1 scores in round two.

2.3 | Representation learning

Having established the accurate GT labels, we moved to CHC classification and phenotyping (Figure 1D). In short, our CG method measures the background signal to estimate the true positive signal and identifies positive cells by gating the fraction of positive pixels within the cell area (**Methods**). While useful for finding potential CHCs, this approach is limited to the cells' intensity profiles and does not take into account spatial localization of protein expression, texture and morphology [42]. Consequently, we applied a β-VAE to the single-cell images, which successfully learned single-cell representations that effectively differentiated CHCs from other PBMCs. We subsequently trained an ML classifier on these representations to improve CHC detection in PBMC samples over CG.

In the uniform manifold approximation and projection [43] (UMAP) visualization of the β-VAE embeddings (Figure 3A,B), the CG labels (color-coded in Figure 3A) showed lower class separation, with more positive cells appearing in the dominantly negative region. However, the GT labels (color-coded in Figure 3B) demonstrate improved separation between classes in the projection. In the class density plots of the CG-colored UMAP (inset in Figure 3A), substantial overlap among the classes was evident. In contrast, when stratified by annotator vote (Figure 3C), the same density plots maintained the positive/negative spatial distribution while showing clearer separation between high-confidence non-CHCs (0 out of 10 votes) and high-confidence CHCs (10 out of 10 votes). This underscores the use of representation learning to achieve clear class separations and capture annotator consensus compared to CG. While some level of patient-induced batch effect was expected, the CHC region generally exhibited a well-mixed
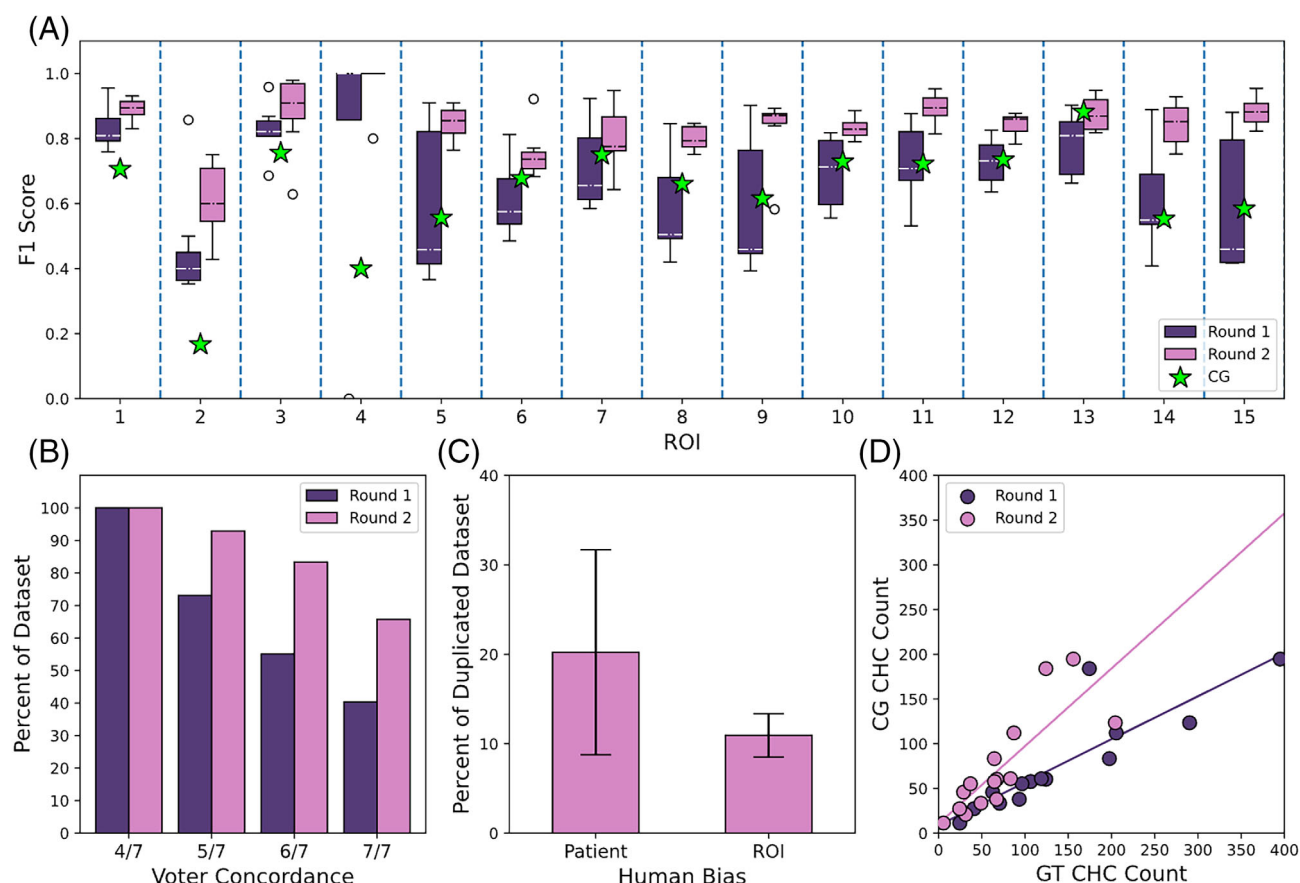
FIGURE 2 Human annotator variation, bias, and alignment with CG. (A) F1 score box-and-whisker plots for Round 1 and Round 2 per ROI. Circles mark outliers and the stars mark the CG score. (B) Inter-annotator agreement across rounds, where each bar is the fraction of cells in the filtered data set that received greater than or equal to the number of votes. (C) Patient and scene bias of seven annotators who completed Round 1 based on the duplicated data set (Methods), with error bars representing ±1 standard deviation. (D) CG CHC count plotted against the human-generated GT counts for Rounds 1 and 2, with corresponding linear regression lines. [Color figure can be viewed at wileyonlinelibrary.com]

patient distribution (Figure S4A). Importantly, the observed CHC enrichment was driven by biologically meaningful features, such as epithelial cocktail and CD45 expression (Figure S4B).

Furthermore, the single-cell representations obtained from the β-VAE exhibit the ability to differentiate cell subtypes characterized by distinct morphologies and staining patterns. For example, six representative cell clusters are shown in Figure 3A (negative CHC region) and Figure 3B (positive CHC region), each displaying unique staining patterns and morphologies. The full UMAP of single-cell images can be seen at https://heussner.github.io/pbmc-umap/ for closer inspection. Notably, the β-VAE's ability to learn morphological features enables it to identify irregularly-shaped cells, which is valuable for detecting poorly segmented cells.

2.4 | PBMC classification

We trained an SVM, a linear model, and a neural network with the single-cell embeddings to classify them as CHCs or non-CHCs. We found that the SVM outperformed the other models in this task. We then compared its accuracy to that of the round one annotators,

round two annotators, and CG, shown in Table 1. The SVM achieved near human-level accuracy on CHC detection, with a fivefold cross-validation average F1 score of 0.80 and a top F1 score of 0.84. By comparison, the round two annotators averaged an F1 score of 0.83 with a top F1 score of 0.84. A paired t test on the fivefold F1 scores revealed a t-statistic of 2.18 (0.095), finding no statistically significant difference between the humans and the SVM. The ROI-induced bias shown in Figure 2C provides additional support for the SVM's ability to match human accuracy, and the SVM outperformed the untrained annotators and CG by a substantial margin.

3 | DISCUSSION

Defining the GT labels for training ML models in biomedical applications is a well-known challenge [44]. While our labels are the product of 10 humans' annotations over multiple iterations, we acknowledge that there are inherent limitations to our approach. Notably, the mosaic single-cell images (**Methods**) provided for annotation differ from native cell images, where cells sit adjacently on the slide, and thus potentially introduce artificial biases into the annotation process.
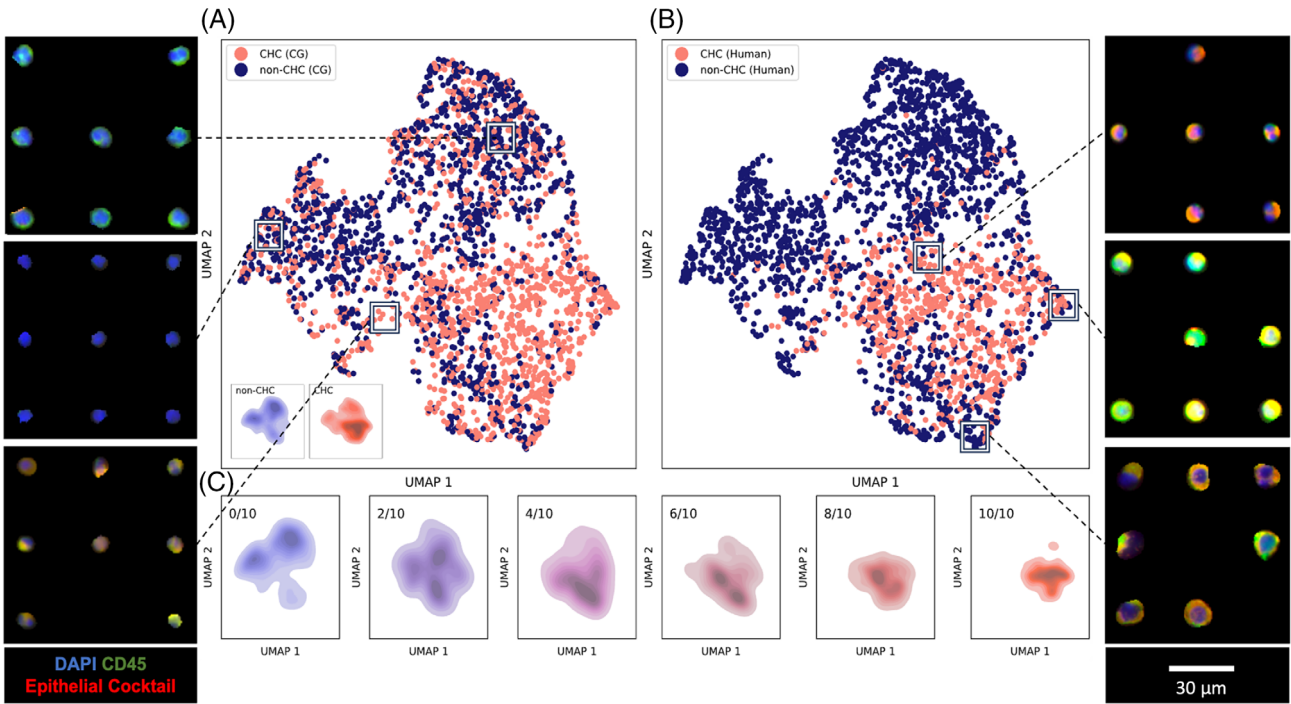
FIGURE 3 UMAP of single-cell β-VAE embeddings. (A) UMAP of β-VAE embeddings from the filtered CHC data set, colored by CG callouts with positive (salmon) and negative (midnight blue) density plots in the lower left corner. (B) UMAP of VAE embeddings from the filtered CHC data set colored by the human GT callouts. (C) Density plots of annotator agreement, ranging from cells with 0 votes (confident negative) to 10 votes (confident positive). Additionally, single-cell images plotted in the UMAP space show clusters of morphologically distinct cells. To view the full single-cell image UMAP, visit https://heussner.github.io/pbmc-umap/. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 PBMC classification summary.

Method	Average F1 (SD)	Top F1
Human Round 1	0.60 (0.010)	0.61
Human Round 2	0.83 (0.0070)	0.84
CG	0.61 (0.023)	0.63
SVM	0.80 (0.027)	0.84

Note: PBMC classification performance summary, reporting average and top F1 scores of a fivefold cross-validation. Human F1 scores were averaged across the participating annotators for each fold, which were then averaged across the fivefold. The CG and SVM F1 scores were averaged across the same fivefold for a similar comparison.

To investigate this possibility, we acquired conventional annotations for each ROI from a single annotator. We then compared these annotations with the GT counts and the CG counts (Figure S5). Our observations revealed a mixed pattern: Some ROIs exhibited higher conventional counts compared to the GT, while others exhibited lower counts. Upon further inspection, we found a strong negative correlation between the difference in conventional count and GT counts with the size of ROI, as shown in Figure 4. ROIs containing over ~50,000 PBMCs show lower conventional counts than GT counts. Conversely, ROIs with fewer than 50,000 PBMC ROIs demonstrated higher conventional counts relative to the GT. We posit that this trend suggests an ROI size-related bias inherent in the traditional
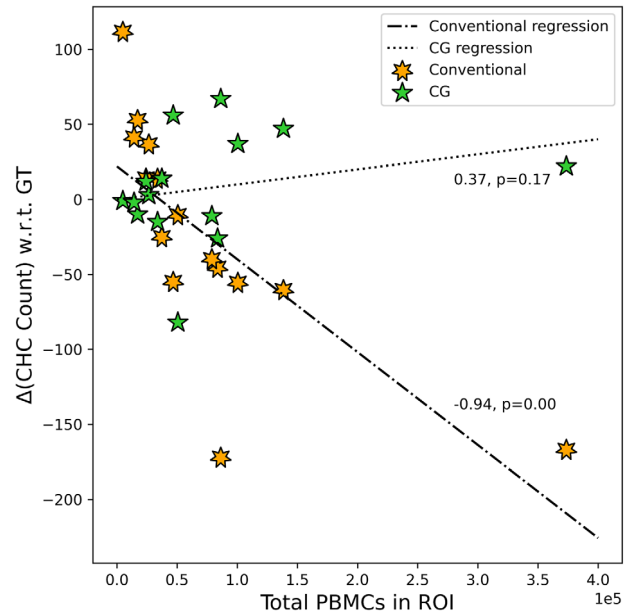

FIGURE 4 Conventional and CG CHC counts with respect to GT count versus ROI PBMC count. Differential CHC count of the conventional method and CG with respect to the GT counts, where $+\Delta$ is a higher count than GT and $-\Delta$ is a lower one. The dotted and dashed-dotted lines are linear regressions, and the text associated with each line is Spearman's rank correlation coefficient and *p*-value. [Color figure can be viewed at wileyonlinelibrary.com]

cell enumeration approach, highlighting the potential oversight of certain CHC populations during the labor-intensive inspection of individual cells within large ROIs.

Another potential limitation of the study was the execution of the second round of annotations. In this round, a single annotator was responsible for setting contrast limits for eight out of the remaining nine annotators, thereby standardizing the epithelial cocktail and CD45 thresholds. However, it is noteworthy that annotator G (Figure S3) deviated from these thresholds. Despite this, annotator G exhibited a comparable reduction in CHC count and an increase in F1 score between rounds one and two, indicating that the review session helped to mitigate variability.

The amount of variation observed in CHC count across seven annotators indicates the potential pitfalls of relying on a single annotator for CHC identification. Such reliance could produce misleading results and affect downstream analyses that are aimed at predicting patient outcome [10]. Understandably, the task of labeling data is expensive and involving multiple annotators might not always be feasible. However, our work highlights that annotator discrepancies largely decrease after training.

The annotation review session allowed the annotators to engage in constructive discussions concerning the staining intensity and morphological characteristics of CHCs. Detailed in the **Methods**, the qualitative criteria adopted for CHC identification encompass various aspects, including the presence of artifacts, shape factors such as eccentricity and size, and the fraction of positive staining signals. The development of these guidelines resulted in enhanced consistency in the GT annotations. Notably, annotators who initially deviated significantly, such as annotator D or annotator G (Figure S3), were able to bring their annotations into better alignment with those of their peers. The annotation study provided this harmonization in a high-throughput and reproducible manner. Therefore, we advocate for the incorporation of similar opportunities for training in future studies that rely on human annotators. This practice could enhance the reliability and validity of annotations, ultimately contributing to more robust research outcomes and greater reliability for clinical readouts.

The image processing pipeline has technical limitations that should be addressed in future work. For example, immunofluorescent contaminants of relatively high intensity can be erroneously identified as cells by segmentation algorithms. These artifacts lack a corresponding nucleus, and a simple approach involves matching the nucleus mask to the cell mask to eliminate those instances. However, occasionally the artifacts overlap true cells. To address these cases, the image was binarized at the estimated positive threshold, and outlier areas (greater than 99% of measured cell areas) were reduced to the estimated background level (**Methods**). Substantial overlap with true cells can occur, in which case the cell cannot be recovered for downstream analysis. Exceptionally large cells, such as cancer-associated macrophage-like cells (CAMLs), can be similar in size to the artifacts and are filtered out with this quality control step. In this study, we focus on CHC detection and not on cells displaying CAML features. Thus, future work should strive for a better balance between artifact management and biological sensitivity.

Additionally, poor segmentation resulted in irregularly shaped cell areas. Aside from size, we did not exclude cells based on their morphology, aiming to encompass a broad spectrum of biological phenotypes, resulting in the inclusion of these imposter cells. For example, Figure S6A shows an imposter cell in row five, column one of the mosaic image. For the biologically informed annotators, these imposters were easily distinguished and not marked as false positives. However, avoiding the cells in the annotation study would have made space for more potential CHCs, improving the GT labels. Moreover, in the case where cells are very close to each other, it would be useful to implement signal correction to account for lateral spillover [45].

To assess the pipeline's generalizability to new patient samples, we performed a ninefold cross-validation, split by patient (Figure S7). The resulting average F1 score of 0.74 was low compared to the previous fivefold cross-validation. The SVM performed worst on patient P2 (ROIs 2–4) with a score of 0.54 and best on patient P6 (ROI 9) with a score of 0.88. However, achieving high accuracy on held-out patients while only training on 8 patients is unreasonable and expected to improve with a larger, more representative data set.

The image analysis pipeline was designed to support multiplexed IF imaging, thereby empowering researchers to explore clinically relevant CHC subtypes by leveraging 20+ biomarkers in combination with morphological and staining pattern differentiation [46]. This approach offers a more comprehensive and nuanced understanding compared to measuring the mean signal intensity of biomarkers. Moreover, β-VAE's capability to learn staining patterns will prove especially beneficial for CAMLs, which typically exhibit punctate epithelial cocktail staining as well as expression of CD45 [47]. However, in this study, which focuses on CHC identification, abnormally large cells were filtered out as a quality control step.

In summary, we combine CG with representation learning to detect CHCs in IF images, achieving near human-level accuracy. We demonstrate the poor agreement between individual annotators and reveal high levels of bias in the annotation process. These observations call for future IF-based cell phenotyping studies to address these concerns. Our detection method and GT generation strategy may be applicable to many rare cell types and are not exclusive to serum samples. Our pipeline enables CHC research in a high-throughput, repeatable manner.

4 | METHODS

4.1 | Human samples

All human specimen studies were approved by the Oregon Health & Science University (OHSU) Institutional Review Board under IRB protocol 5169. Informed consent was obtained from all subjects under IRB protocols for the Oregon Colorectal Cancer Registry, the OHSU Biolibrary, and the Oregon Pancreatic Tumor Registry. Peripheral blood samples were obtained from patients treated at OHSU (Table S1).

4.2 | Sample preparation, staining, and imaging

Patient peripheral blood was collected in heparin-coated vacutainer tubes and processed within 2 h using standard approaches for PBMC isolation [11, 12]. Briefly, whole blood was diluted with phosphate-buffered saline (PBS) and then subjected to density centrifugation with Ficoll-Paque PLUS (GE Healthcare, 17–1440-03) as previously described [21]. After isolation, PBMCs were spread onto glass slides coated with Poly-D Lysine (1 mg/mL, Millipore Sigma, A-003-E) at a density of 600,000 cells per slide, and allowed to dry for 15–30 min. Cells were then fixed for 20 min with 4% paraformaldehyde, washed with PBS 3 × 2 min, and subjected to increasing ethanol baths 70%, 90%, and 100% ethanol. Slides were air-dried at room temperature, and stored at 4°C. Antibody staining was performed by rehydrating cell slides, then blocking non-specific epitopes with BlockAid™ Blocking Solution (ThermoFisher). To identify CHCs, PBMCs were incubated with fluorescent-conjugated antibodies against epithelial proteins ECAD, EpCAM, panCK, and the pan-leukocyte antibody CD45 (Table S2). Antibodies were diluted in BlockAid (Thermo Fisher, Waltham, MA) at 22°C for 45 mins or at 4°C for 16 h. The antibody panel was validated in our previously published studies [10–12, 48]. Cells were counterstained with DAPI and coverslipped using fluoromount G (Invitrogen, 00-4958-02). Stained specimens were digitally scanned at ×20 magnification with a ZEISS AxioScanner. Z1 with a Colibri 7 light source (Zeiss) equipped with the following filter cubes DAPI (Semrock, LED-DAPI-A-000), AF488 (Zeiss 38 HE), AF555 (Zeiss 43 HE), AF647 (Zeiss 50), and AF750 (Chroma 49,007 ET Cy7). ROIs that avoided scanning and staining artifacts were selected and exported using ZEISS ZEN blue software version 3.4 (Carl ZEISS AG, Oberkochen, Germany).

4.3 | Image preprocessing

We utilized MCMICRO [49] to stitch (ASHLAR) [50] and correct background illumination (BaSiC) [51]. Next, we employed DeepCell's Mesmer [40] to segment the nuclei with the DAPI channel and cells with the projection of epithelial cocktail and CD45 channels.

4.4 | Quality control

One unforeseen challenge encountered with serum samples was the presence of irregular cell size distributions (Figure S8A). To address this, we first developed a visualization tool to generate heatmaps plotting DAPI intensity, cell major axis length, and cell area on the original slide image (Figure S8B). These heatmaps allow us to easily evaluate the adequacy of the blood smear protocol, staining process and image acquisition. To remove imaging artifacts (Figure S8C), we filter out cells above the 99th percentile and below the 1st percentile of cell size (Figure S8D). We also excluded cells lacking corresponding nuclei to ensure data integrity.

4.5 | Computational gating

We gated the epithelial cocktail and CD45 by first estimating their background levels by measuring the mean (μ) and standard deviation (σ) of their corresponding image channel's background intensity. Then, we set a positive threshold by adding 3σ to μ. We multiplied the epithelial cocktail threshold by an adjustment factor of 1.5 to account for a low signal-to-noise ratio relative to CD45. Next, we compute the positive pixel ratio (PPR) for each cell, which represents the fraction of positive pixels within the cell area. The PPR values were then thresholded using an elbow plot optimization approach. We confirmed the reproducibility of this strategy by comparing elbow curves for each ROI, as shown in Figure S9. These PPR thresholds were crucial in determining whether a cell was positive or negative for a given marker. While this step is tailored to CD45 and epithelial cocktail, it exhibited broad applicability and could be extended to other biomarkers for serum samples.

4.6 | Annotation study design

To carry out the annotation study, we curated a single-cell mosaic image containing N CG epithelial cocktail$^+$/CD45$^+$ cells from the previous step and the next N cells below the PPR threshold of epithelial cocktail$^+$ ("potential-CHCs") for each ROI. We chose epithelial cocktail as the limiting biomarker for selecting potential CHCs due to its relative rarity compared to CD45; We expected more than 95% of PBMCs to be CD45$^+$, while less than 1% of PBMCs to be epithelial cocktail$^+$. We also added N negative control cells which were determined by an epithelial cocktail PPR of zero and CD45 PPR below the previously determined positive threshold. Negative controls comprised the top row of the mosaic, which the annotators used to set their histogram thresholds. To study annotator bias, we created a bank of CG CHCs and duplicated them into other ROI mosaics. Specifically, we duplicated cells and placed them in neighboring ROIs from the same patient sample (98 cells), and placed others in ROIs from different patient samples (746 cells). These duplicates comprised 20% of the total cells in each mosaic. We added cells to the mosaics by cropping them from the processed ROI and padding them to be 96 by 96 pixels. Each mosaic was a 16-bit RGBA image; epithelial cocktail (Red), CD45 (Green), DAPI (Blue), and gridlines (Alpha). We trained annotators to use ImageJ [32] software to annotate the mosaic cells as CHCs or non-CHCs with the paint tool. An example mosaic side-by-side with human annotations is shown in Figure S6.

4.7 | Histogram settings for CHC annotation

Visualization settings were adjusted for each ROI based on negative control cells. Specifically, the minimum and maximum values were set to just below what could be visibly detected in the negative control cells along the top row of the mosaic for both CD45 and epithelial

cocktail channels. To account for differences in CHC calling due to histogram setting variation between annotators, the minimum and maximum values were set by a single annotator and distributed to other annotators before their annotation in round two. Annotator G did not follow this protocol, instead setting their histogram values.

4.8 | CHC annotation consensus

After completing the first round of CHC annotations, all annotators met to review the data and determine a set of consensus features for CHC calling. The set of criteria that were agreed upon includes the following: (1) Cells must have an intact nucleus with adequate DAPI signal (Figure S10A); (2) Signal distribution of CD45 and epithelial cocktail should be around the majority of the cell membrane/cytoplasm (i.e., cells with bright specks of signal should not be called) (Figure S10B); (3) Cells with very dim and diffuse CD45 or epithelial cocktail signal should not be called (Figure S10C). The criteria were distributed to three additional annotators before completing the second round of annotations.

4.9 | CHC enumeration

All reported CHC counts were normalized to 50,000 PBMCs to fairly compare ROIs with varying image sizes and cell densities [10].

4.10 | β-variational autoencoder

We trained the β-VAE [52] on the potential-CHCs, excluding the negative control cells and cell duplicates used in the annotation study, for a total of 2834 cells. Our β-VAE had a convolutional encoder with five layers of 32, 64, 128, 256, and 512 filters and a mirrored decoder, with a latent dimension of 64. The single-cell images were resized to 128 by 128 pixels and rescaled to an intensity range of [0,1]. We employed a β scheduler during training, starting at $\beta = 0.001$ and increasing in intervals of 0.001 every batch until $\beta = 2$ was reached and remained until training was complete [53]. The batch size was four and the learning rate was 0.0005, with early stopping and a maximum number of epochs set to 1000. The model was built with the PyTorch Lightning platform and trained on one NVIDIA A40 GPU.

4.11 | SVM classification

We employed an SVM as the PBMC classifier in our study. We chose a second-degree polynomial kernel with a coefficient of one. The regularization parameter was set to one, and the gamma parameter was also set to one. To assess the SVM's performance, we conducted a fivefold cross-validation. This approach partitions the data set into five subsets, using four of them for training and one for testing in each iteration. The process is repeated five times, and the final evaluation is obtained by averaging the results, providing a robust estimate of the model's generalizability. By employing the SVM with the specified parameters, we achieved the highest F1 score compared with a linear model and neural network.

CONFLICT OF INTEREST STATEMENT
The authors declare no competing interests.

DATA AVAILABILITY STATEMENT
All single-cell images used in the annotation study are publicly available at https://zenodo.org/record/8270693. All software used in this manuscript is detailed in the article's Methods section and its Supplemental Information. The associated scripts are freely available via GitHub as described at https://github.com/heussner/crc-process.

ORCID
Robert T. Heussner https://orcid.org/0009-0005-7690-1934
Young Hwan Chang https://orcid.org/0000-0001-8764-1959

REFERENCES
1. Cao W, Chen H-D, Yu Y-W, Li N, Chen W-Q. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. Chin Med J. 2021;134:783–91.
2. Hirahata T, ul Quraish R, Quraish A, ul Quraish S, Naz M, Razzaq MA. Liquid biopsy: a distinctive approach to the diagnosis and prognosis of cancer. Cancer Informat. 2022;21:11769351221076062.

3. Song P, Wu LR, Yan YH, Zhang JX, Chu T, Kwong LN, et al. Limitations and opportunities of technologies for the analysis of cell-free DNA in cancer diagnostics. Nat Biomed Eng. 2022;6:232–45.

4. Gao Y, Wang J, Zhou Y, Sheng S, Qian SY, Huo X. Evaluation of serum CEA, CA19-9, CA72-4, CA125 and ferritin as diagnostic markers and factors of clinical parameters for colorectal cancer. Sci Rep. 2018;8:2732.

5. Sutton TL, Patel RK, Anderson AN, Bowden SG, Whalen R, Giske NR, et al. Circulating cells with macrophage-like characteristics in cancer: the importance of circulating neoplastic-immune hybrid cells in cancer. Cancer. 2022;14:3871.

6. Cohen SJ, Punt CJA, Iannotti N, Saidman BH, Sabbath KD, Gabrail NY, et al. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. J Clin Oncol. 2008;26:3213–21.

7. Bidard F-C, Jacot W, Kiavue N, Dureau S, Kadi A, Brain E, et al. Efficacy of circulating tumor cell count-driven vs clinician-driven first-line therapy choice in hormone receptor-positive, ERBB2-negative metastatic breast cancer: the STIC CTC randomized clinical trial. JAMA Oncol. 2021;7:34–41.

8. Cabel L, Berger F, Cottu P, Loirat D, Rampanou A, Brain E, et al. Clinical utility of circulating tumour cell-based monitoring of late-line chemotherapy for metastatic breast cancer: the randomised CirCe01 trial. Br J Cancer. 2021;124:1207–13.

9. Smerage JB, Barlow WE, Hortobagyi GN, Winer EP, Leyland-Jones B, Srkalovic G, et al. Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. J Clin Oncol. 2014;32:3483–9.

10. Gast CE, Silk AD, Zarour L, Riegler L, Burkhart JG, Gustafson KT, et al. Cell fusion potentiates tumor heterogeneity and reveals circulating hybrid cells that correlate with stage and survival. Sci Adv. 2018;4:eaat7828.

11. Walker BS, Sutton TL, Zarour L, Hunter JG, Wood SG, Tsikitis VL, et al. Circulating hybrid cells: a novel liquid biomarker of treatment response in gastrointestinal cancers. Ann Surg Oncol. 2021;28:8567–78.

12. Dietz MS, Sutton TL, Walker BS, Gast CE, Zarour L, Sengupta SK, et al. Relevance of circulating hybrid cells as a non-invasive biomarker for myriad solid tumors. Sci Rep. 2021;11:13630.

13. Ramakrishnan M, Mathur SR, Mukhopadhyay A. Fusion-derived epithelial cancer cells express hematopoietic markers and contribute to stem cell and migratory phenotype in ovarian carcinoma. Cancer Res. 2013;73:5360–70.

14. Lazova R, LaBerge GS, Duvall E, Spoelstra N, Klump V, Sznol M, et al. A melanoma Brain metastasis with a donor-patient hybrid genome following bone marrow transplantation: first evidence for fusion in human cancer. PLoS One. 2013;8:e66731.

15. LaBerge GS, Duvall E, Grasmick Z, Haedicke K, Pawelek J. A melanoma lymph node metastasis with a donor-patient hybrid genome following bone marrow transplantation: a second case of leucocyte-tumor cell hybridization in cancer metastasis. PLoS One. 2017;12:e0168581.

16. Cao M-F, Chen L, Dang WQ, Zhang XC, Zhang X, Shi Y, et al. Hybrids by tumor-associated macrophages × glioblastoma cells entail nuclear reprogramming and glioblastoma invasion. Cancer Lett. 2019;442:445–52.

17. Yilmaz Y, Lazova R, Qumsiyeh M, Cooper D, Pawelek J. Donor Y chromosome in renal carcinoma cells of a female BMT recipient: visualization of putative BMT-tumor hybrids by FISH. Bone Marrow Transplant. 2005;35:1021–4.

18. Xu M-H, Gao X, Luo D, Zhou XD, Xiong W, Liu GX. EMT and acquisition of stem cell-like properties are involved in spontaneous formation of tumorigenic hybrids between lung cancer and bone marrow-derived mesenchymal stem cells. PLoS One. 2014;9:e87893.

19. Luo F, Liu T, Wang J, Li J, Ma P, Ding H, et al. Bone marrow mesenchymal stem cells participate in prostate carcinogenesis and promote growth of prostate cancer by cell fusion in vivo. Oncotarget. 2016;7:30924–34.

20. Laberge GS, Duvall E, Haedicke K, Pawelek J. Leukocyte⁻cancer cell fusion-genesis of a deadly journey. Cell. 2019;8:170.

21. Parappilly MS, Chin Y, Whalen RM, Anderson AN, Robinson TS, Strgar L, et al. Circulating neoplastic-immune hybrid cells predict metastatic progression in uveal melanoma. Cancer. 2022;14:4617.

22. Wang L, Balasubramanian P, Chen AP, Kummar S, Evrard YA, Kinders RJ. Promise and limits of the CellSearch platform for evaluating pharmacodynamics in circulating tumor cells. Semin Oncol. 2016;43:464–75.

23. Sollier E, Go DE, Che J, Gossett DR, O'Byrne S, Weaver WM, et al. Size-selective collection of circulating tumor cells using vortex technology. Lab Chip. 2014;14:63–77.

24. Di Trapani M, Manaresi N, Medoro G. DEPArray™ system: an automatic image-based sorter for isolation of pure circulating tumor cells. Cytometry A. 2018;93:1260–6.

25. Sutton TL, Walker BS, Wong MH. Circulating hybrid cells join the fray of circulating cellular biomarkers. Cell Mol Gastroenterol Hepatol. 2019;8:595–607.

26. Lin J-R, Fallahi-Sichani M, Sorger PK. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. Nat Commun. 2015;6:8390.

27. Burlingame EA, Eng J, Thibault G, Chin K, Gray JW, Chang YH. Toward reproducible, scalable, and robust data analysis across multiplex tissue imaging platforms. Cell Rep Methods. 2021;1:100053.

28. Chawla NV, Japkowicz N, Ko A. Editorial: special issue on learning from imbalanced data sets. SIGKDD Explor Newsl. 2004;6:1–6.

29. Lazcano R, Rojas F, Laberiano C, Hernandez S, Parra ER. Pathology quality control for multiplex immunofluorescence and image analysis assessment in longitudinal studies. Front Mol Biosci. 2021;8:661222.

30. Chang YH, Chin K, Thibault G, Eng J, Burlingame E, Gray JW. RESTORE: robust intEnSiTy nORmalization mEthod for multiplexed imaging. Commun Biol. 2020;3:111.

31. Harris CR, McKinley ET, Roland JT, Liu Q, Shrubsole MJ, Lau KS, et al. Quantifying and correcting slide-to-slide variation in multiplexed immunofluorescence images. Bioinformatics. 2022;38:1700–7.

32. Collins TJ. ImageJ for microscopy. Biotechniques. 2007;43:25–30.

33. Svensson C-M, Hübler R, Figge MT. Automated classification of circulating tumor cells and the impact of Interobsever variability on classifier training and performance. J Immunol Res. 2015;2015:573165.

34. Saucedo-Zeni N, Mewes S, Niestroj R, Gasiorowski L, Murawa D, Nowaczyk P, et al. A novel method for the in vivo isolation of circulating tumor cells from peripheral blood of cancer patients using a functionalized and structured medical wire. Int J Oncol. 2012;41:1241–50.

35. Poellmann MJ, Bu J, Liu S, Wang AZ, Seyedin SN, Chandrasekharan C, et al. Nanotechnology and machine learning enable circulating tumor cells as a reliable biomarker for radiotherapy responses of gastrointestinal cancer patients. Biosens Bioelectron. 2023;226:115117.

36. Zeune LL, Boink YE, van Dalum G, Nanou A, de Wit S, Andree KC, et al. Deep learning of circulating tumour cells. Nat Mach Intell. 2020;2:124–33.

37. Shen C, Rawal S, Brown R, Zhou H, Agarwal A, Watson MA, et al. Automatic detection of circulating tumor cells and cancer associated fibroblasts using deep learning. Sci Rep. 2023;13:5708.

38. Guo Z, Lin X, Hui Y, Wang J, Zhang Q, Kong F. Circulating tumor cell identification based on deep learning. Front Oncol. 2022;12:843879.

39. Cheng KS, Pan R, Pan H, Li B, Meena SS, Xing H, et al. ALICE: a hybrid AI paradigm with enhanced connectivity and cybersecurity for a serendipitous encounter with circulating hybrid cells. Theranostics. 2020;10:11026–48.

40. Greenwald NF, Miller G, Moen E, Kong A, Kagel A, Dougherty T, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat Biotechnol. 2022;40:555–65.

41. Henn TE, Anderson AN, Hollett YR, Sutton TL, Walker BS, Swain JR, et al. Circulating hybrid cells predict presence of occult nodal metastases in oral cavity carcinoma. Head Neck. 2021;43:2193–201.

42. Ternes L, Dane M, Gross S, Labrie M, Mills G, Gray J, et al. A multi-encoder variational autoencoder controls multiple transformational features in single-cell image analysis. Commun Biol. 2022;5:255.

43. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. 2018.

44. Lebovitz S, Levina N, Lifshitz-Assaf H. Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. MIS Q. 2021;45:1501–26.

45. Bai Y, Zhu B, Rovira-Clave X, Chen H, Markovic M, Chan CN, et al. Adjacent cell marker lateral spillover compensation and reinforcement for multiplexed images. Front Immunol. 2021;12:652631.

46. Einhaus J, Rochwarger A, Mattern S, Gaudillière B, Schürch CM. High-multiplex tissue imaging in routine pathology-are we there yet? Virchows Arch. 2023;482:801–12.

47. Adams DL, Martin SS, Alpaugh RK, Charpentier M, Tsai S, Bergan RC, et al. Circulating giant macrophages as a potential biomarker of solid tumors. Proc Natl Acad Sci U S A. 2014;111:3514–9.

48. Powell AE, Anderson EC, Davies PS, Silk AD, Pelz C, Impey S, et al. Fusion between intestinal epithelial cells and macrophages in a cancer context results in nuclear reprogramming. Cancer Res. 2011;71: 1497–505.

49. Schapiro D, Sokolov A, Yapp C, Chen YA, Muhlich JL, Hess J, et al. MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging. Nat Methods. 2022;19:311–5.

50. Muhlich JL, Chen YA, Yapp C, Russell D, Santagata S, Sorger PK. Stitching and registering highly multiplexed whole-slide images of tissues and tumors using ASHLAR. Bioinformatics. 2022;38:4613–21.

51. Peng T, Thorn K, Schroeder T, Wang L, Theis FJ, Marr C, et al. A BaSiC tool for background and shading correction of optical microscopy images. Nat Commun. 2017;8:14836.

52. Higgins I, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, et al. Beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. 2016.

53. Fu H, Li C, Liu X, Gao J, Asli C, Carin L. Cyclical Annealing Schedule: a Simple Approach to Mitigating KL Vanishing. 2019.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.