Cancer Research

Platform for Quantitative Evaluation of Spatial Intratumoral Heterogeneity in Multiplexed Fluorescence Images

Daniel M. Spagnolo[1,2], Yousef Al-Kofahi[3], Peihong Zhu[3], Timothy R. Lezon[2,4], Albert Gough[2,4], Andrew M. Stern[2,4], Adrian V. Lee[5,6], Fiona Ginty[7], Brion Sarachan[3], D. Lansing Taylor[2,4,5], and S. Chakra Chennubhotla[2]

Abstract

We introduce THRIVE (Tumor Heterogeneity Research Interactive Visualization Environment), an open-source tool developed to assist cancer researchers in interactive hypothesis testing. The focus of this tool is to quantify spatial intratumoral heterogeneity (ITH), and the interactions between different cell phenotypes and noncellular constituents. Specifically, we foresee applications in phenotyping cells within tumor microenvironments, recognizing tumor boundaries, identifying degrees of immune infiltration and epithelial/stromal separation, and identification of heterotypic signaling networks underlying microdomains. The THRIVE platform provides an integrated workflow for analyzing whole-slide immunofluorescence images and tissue microarrays, including algorithms for segmentation, quantification, and heterogeneity analysis. THRIVE promotes flexible deployment, a maintainable code base using open-source libraries, and an extensible framework for customizing algorithms with ease. THRIVE was designed with highly multiplexed immunofluorescence images in mind, and, by providing a platform to efficiently analyze high-dimensional immunofluorescence signals, we hope to advance these data toward mainstream adoption in cancer research. *Cancer Res; 77(21); e71–74. ©2017 AACR.*

Introduction

Spatial intratumoral heterogeneity (ITH), quantified as the number and variation of cell phenotypes, as well as the spatial relationships between cells and extracellular molecules within a tumor microenvironment (TME), is of high prognostic and diagnostic value (1–3). The acknowledgement of spatial ITH as a key factor in tumor progression has identified a need for new informatics tools to quantify spatial heterogeneity in cancer research applications.

Toward this end, we have created an open-source tool, THRIVE (Tumor Heterogeneity Research Interactive Visualization Environment), which (i) permits visualization of large cohorts of whole-slide images and tissue microarrays; (ii) performs interactive image analysis tasks such as cell segmentation, cell phenotyp- ing, and tumor microdomain discovery via ITH; and (iii) contains statistical inference tools to aid in cancer-specific hypothesis testing. We adopt the term tumor microdomain to describe phenotypically distinct regions of the TME, which represent a fundamental unit of spatial heterogeneity (4). This software platform encapsulates a workflow for quantifying ITH in immunofluorescence (IF) images ranging from a single biomarker to standard multiplexed biomarkers (up to 7) to emerging hyperplexed (>7) images (5, 6). Each additional biomarker in IF images allows for more insight into cellular and disease mechanisms, but increases cost and data acquisition complexity, so it was important to develop a platform applicable to a range of imaging modalities. Existing image analysis tools, such as CellProfiler (7), ImageJ/Fiji (8), and BioimageXD (9), although useful, are very general tools and thus contain only several of the required features necessary for analyzing spatial ITH, especially from multiplexed and hyperplexed IF images. Although some of these contain colocalization pipelines for measuring spatial coincidence of biomarkers within single cells, THRIVE incorporates novel information theoretic measures (pointwise mutual information) and current ecological diversity metrics (quadratic entropy) to enhance insights into the spatial organization of tumors by looking at interactions between cells in the TME (1, 10). We provide the added benefit of designing algorithms with high dimensional image data in mind, collected through multiplexed IF, mass spectrometry, or other data collection methods that allow for a large array of molecular probes. THRIVE allows for the creation of custom workflows with plug-in architecture for new functions, can potentially link to genomic and clinical data, and provides multiple spatial- and population-based heterogeneity metrics for ease of use by cancer biologists and clinicians alike.

[1]Program in Computational Biology, Joint Carnegie Mellon University-University of Pittsburgh, Pittsburgh, Pennsylvania. [2]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania. [3]Software Science and Analytics Organization, GE Global Research Center, Niskayuna, New York. [4]Drug Discovery Institute, University of Pittsburgh, Pittsburgh, Pennsylvania. [5]University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania. [6]Department of Pharmacology and Chemical Biology, University of Pittsburgh, Pittsburgh, Pennsylvania. [7]Biosciences Organization, GE Global Research Center, Niskayuna, New York.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Author: Chakra Chennubhotla, University of Pittsburgh, BST 3 6th Floor, Pittsburgh, PA 15213. Phone: 412-648-7794; Fax: 412-648-3163; E-mail: chakracs@pitt.edu

doi: 10.1158/0008-5472.CAN-17-0676

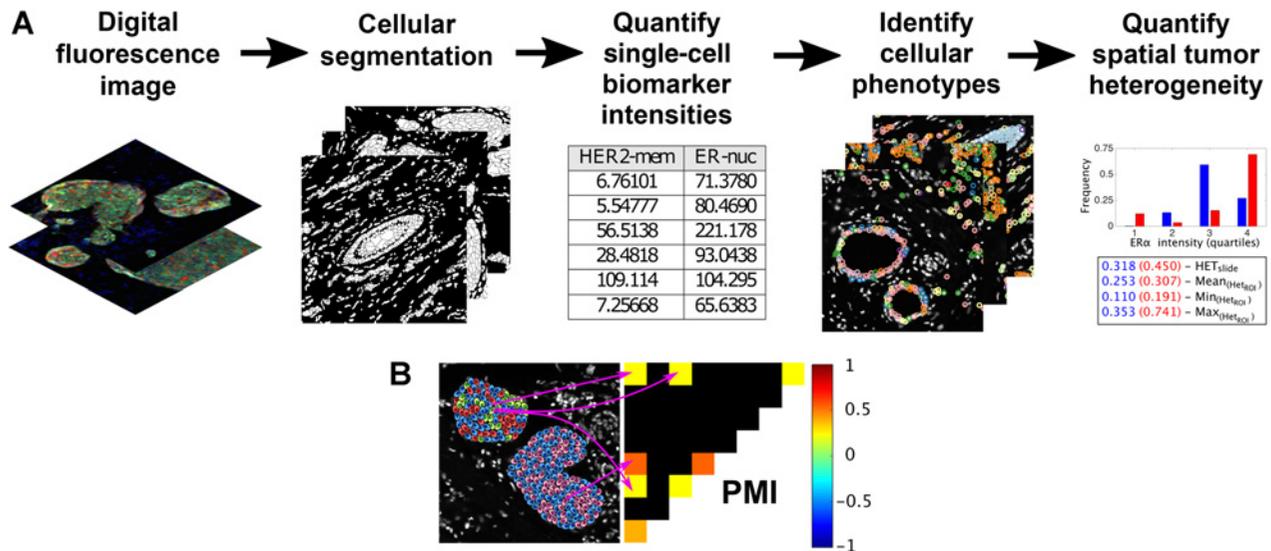©2017 American Association for Cancer Research.

AACR e71

Figure 1.
THRIVE. **A,** For a given panel of images, a cell segmentation algorithm is run to obtain single-cell resolution. Then, biomarker intensity statistics (e.g., mean, median) are computed for each cell from the segmentation results. These statistics are used to discover cell phenotypes via pattern recognition. Heterogeneity metrics are used to quantify the spatial relationships between cell phenotypes. The bar graph shows the heterogeneity of cell phenotypes discovered from ERα expression for two different tumor ROIs (shown in red and blue). Phenotype heterogeneity is quantified by quadratic entropy summarized over the whole slide and statistics from ROIs. **B,** Pointwise mutual information (PMI) maps capture the relative spatial cooccurrences of cell phenotypes (denoted by various cell colors) in a multiplexed IF image (1). The diagonal elements of the pointwise mutual information map denote globally heterogeneous and locally homogenous interactions, while off-diagonal elements capture locally heterogeneous interactions. Pointwise mutual information is scaled from -1 (negative association) to 1 (positive association), where 0 is the background cooccurrence of cell phenotypes.

Thrive Platform Description

A computational cancer researcher will find THRIVE to be (i) extensible such that cancer researchers can easily add new experimental tumor heterogeneity algorithms and datasets; (ii) maintainable within the research community by leveraging existing open-source libraries, therefore minimizing custom code; and (iii) flexible to deploy in a variety of environments, from local research laboratory installations to cloud deployments shared by the research community. THRIVE uses Docker, which has the advantage of consistent behavior and easy deployment on laptops, computer workstations, and on cloud services.

THRIVE's file structure is very general and can easily import files from a variety of microscope platforms. The top level contains each microscope slide directory; the second level stores directories for each imaged regions in a particular slide, and the third level contains both the source image directory and results directories for each region. Under the source image directory are folders for each acquired channel, and under the results directory are folders for each algorithm's (e.g., segmentation, quantification) output.

Ease of integrating new analysis methods has been an overriding requirement in our design of THRIVE. The straightforward steps that an algorithm developer needs to follow are detailed in THRIVE's technical documentation. Briefly, the developer would write a short script that pulls input files from data storage, launches the algorithm, stores result, checks for errors, and returns status information. This script is then packaged with the generic THRIVE code in a Docker container. In addition, the inputs, parameters, outputs, and UI display choices need to be explicitly described, and a Docker Compose file is needed to alert THRIVE that a new algorithm is available. Any programming language can be used to develop heterogeneity algorithms, identified and vetted by the research community, and added to THRIVE. Source code will be available on GitHub, and through the ITCR webpage. THRIVE's project website is located at ith.csb.pitt.edu.

Thrive Capabilities

THRIVE provides a user interface that enables the researcher to browse and review multichannel whole-slide IF images, request single-cell segmentation and quantification, and review results. The researcher can then run a variety of tumor heterogeneity algorithms and review and compare those results (Fig. 1A). Our platform enables any number of alternate segmentation, quantification, and heterogeneity algorithms to be integrated into the image-processing workflow, both algorithms that we plan to include with THRIVE (e.g., single-cell segmentation vs. subcellular-resolution segmentation) and algorithms to be developed and shared by the research community.

THRIVE was developed for the analysis of the spatial distributions of biomarkers, typically using IF labeling, and typically in tissue sections on slides. Other methods such as mass spectrometry imaging could also be used to generate compatible images of multiple biomarkers. The following three classes of imaging systems can generate IF images compatible with THRIVE: commercial slide scanning systems, high content screening (HCS) systems, and general purpose microscopy systems. Commercial slide scanning systems with multichannel fluorescence capability include the PerkinElmer Vectra, Leica

Aperio FL, Hamamatsu Nanozoomer, and others. In general, these are the fastest and most efficient, as they are optimized for slide scanning. Most HCS systems, including the PerkinElmer Opera, Molecular Devices ImageXpress, Thermo Fisher Array-scan, GE INCell, and others, have slide holders and can collect multichannel fluorescence images. HCS systems are also fast and efficient for high volume imaging. General purpose fluorescence microscope systems from Olympus, Nikon, Zeiss, Leica, and others can be used to acquire images from slides, using software packages from the manufacturer, or open-source solutions like Micromanager with the Slide Explorer plug-in. These systems are less efficient, but more cost effective than slide scanners or HCS systems. In all cases, images can be easily saved as TIF files and imported for analysis.

A typical workflow, as demonstrated in Supplementary Video S1, starts with a cell segmentation step, allowing single-channel and two-channel segmentation of individual cells. If only a cell nuclei channel is available, a single-channel segmentation algorithm delineates the individual nuclei in the image (e.g., ref. 11) and then extracts synthetic cell boundary approximations using Voronoi tessellation. When a membrane cell-marker channel is also available, first the cell nuclei are segmented and cell boundary approximations are extracted as described above, and then a watershed algorithm refines the cell boundaries using the additional cell-marker channel data. Each cell is assigned a unique ID, and each subcellular pixel is assigned into one of two compartments: nuclear or extranuclear.

In the biomarker quantification step, cell and subcellular-level statistics (e.g., mean, SD, mode, etc.) are computed for each available biomarker, as are cell morphometric features (location, area, and cell radius). These measurements can be used to calculate the Pittsburgh indices (12) and other measures of cell-level phenotypic heterogeneity. Once each cell is described as a single point in a multivariate feature space, phenotypes can be identified through standard clustering techniques. We define phenotypes in this context as the combinations of biomarkers and expression levels in subpopulations of cells. Currently, THRIVE uses basic k-means clustering for cell phenotyping by finding k groups of similar cells in an N dimensional space (where N is the number of biomarkers in the image) and will soon incorporate k-SVD (13) phenotyping used in our previous work (1). The benefit of k-SVD is that it finds a lower dimensional space with which to group cells into phenotypes and seeks a sparse representation of the cell data where there is less ambiguity about phenotyping cells that border on two potential phenotypes.

Spatial heterogeneity is characterized by microdomains, which we define as subpopulations of cells clustered together, considered not only by the relative populations of the cellular phenotypes within it, but by their spatial distribution. The toolkit contains a starter set of methods to quantify spatial heterogeneity, such as our own technique based on pointwise mutual information (1). Using these methods, users can compare microdomains to one another or to the whole tumor (Fig 1B).

Application in Tumor Heterogeneity

Using the THRIVE cell quantification algorithms, tumor cells, as well as an array of different immune cell types, can be identified. THRIVE can quantify the statistically significant cooccurrences between various cell types within tumor microdomains and at microdomain interfaces, which are often associated with known intratumor phenomena. For example, THRIVE can measure the degree to which epithelial and stromal cells are intermixed or spatially separated (14) and can determine the amount of immune infiltration (i.e., the degree to which immune cells invade the TME) within a tumor sample or region of interest (ROI; ref. 15), both of which have prognostic potential. The predictive power of the spatial relationships between various immune cells and tumor cells can be applied as a cancer biomarker for immune infiltration. In addition, the identification of tumor and nontumor cells can be used to locate microdomains such that the interfaces between dissimilar microdomains can identify tumor boundaries. This tool will be useful in automating ROI discovery and assisting pathologists in a computational pathology digital slide workflow.

Notably, THRIVE can be used to identify microdomains containing spatial clusters of network signatures contributed by oncogenic signaling pathways. For example, in the PI3K pathway, genetic alterations are found in most invasive breast cancers, and PIK3CA mutations are hypothesized to drive carcinogenesis in the breast. Using THRIVE workflows, one can assess the emerging spatial heterogeneity in the PI3K pathway and identify microdomains containing common signatures, for example, the epithelial–stromal interface PI3K/MAPK signature (16). Similar efforts to study the MTOR pathway in colorectal cancer (5) could also be assisted by using THRIVE.

We envision that THRIVE will enable the determination of a mechanistic link between spatial ITH quantification and cancer progression. It has been shown that neoadjuvant chemotherapy for cancer results in changes in spatial heterogeneity that correlate with poor long-term outcome following adjuvant therapy (3). As long-term survival is largely defined by progression to metastatic disease, these results suggest that particular microdomains within the primary tumor impart metastatic potential to a subpopulation of treatment-resistant tumor cells. Implementation of our platform presents a unique opportunity to identify the heterotypic signaling networks within these metastasis-conferring domains that can lead to robust biomarkers mechanistically linked to disease progression and optimized therapeutic strategies for individual patients.

Disclosure of Potential Conflicts of Interest

D.L. Taylor is the chairman at and has ownership interest (including patents) in Spatial Pathology Diagnostics Inc. S.C. Chennubhotla is the president at and has ownership interest (including patents) in Spatial Pathology Diagnostics Inc. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: D.M. Spagnolo, Y. Al-Kofahi, A.M. Stern, A.V. Lee, B. Sarachan, D.L. Taylor, S.C. Chennubhotla
Development of methodology: D.M. Spagnolo, Y. Al-Kofahi, P. Zhu, T.R. Lezon, B. Sarachan, S.C. Chennubhotla
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): F. Ginty
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): D.M. Spagnolo, T.R. Lezon, A.V. Lee, S.C. Chennubhotla
Writing, review, and/or revision of the manuscript: D.M. Spagnolo, Y. Al-Kofahi, P. Zhu, T.R. Lezon, A. Gough, A.M. Stern, B. Sarachan, D.L. Taylor, S.C. Chennubhotla
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): F. Ginty
Study supervision: F. Ginty, D.L. Taylor, S.C. Chennubhotla

References

1. Spagnolo DM, Gyanchandani R, Al-Kofahi Y, Stern AM, Lezon TR, Gough A, et al. Pointwise mutual information quantifies intratumor heterogeneity in tissue sections labeled with multiple fluorescent biomarkers. J Pathol Inform 2016;7:47.
2. Waclaw B, Bozic I, Pittman ME, Hruban RH, Vogelstein B, Nowak MA. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. Nature 2015;525:261–4.
3. Janiszewska M, Liu L, Almendro V, Kuang Y, Paweletz C, Sakr RA, et al. In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. Nat Genet 2015;47: 1212–9.
4. Sipkins DA, Wei X, Wu JW, Runnels JM, Cote D, Means TK, et al. *In vivo* imaging of specialized bone marrow endothelial microdomains for tumour engraftment. Nature 2005;435:969–73.
5. Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, Bordwell A, et al. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proc Natl Acad Sci U S A 2013;110:11982–7.
6. Gough A, Lezon T, Faeder JR, Chennubhotla C, Murphy RF, Critchley-Thorne R, et al. High content analysis with cellular and tissue systems biology: a bridge between cancer cell biology and tissue-based diagnostics. In: Mendelsohn J, Howley PM, Israel MA, Gray JW, Thompson C, editors. The molecular basis of cancer. 4th ed. Philadelphia, PA: Saunders/Elsevier; 2015. p. 369–92.
7. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 2006;7:R100.
8. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 2012;9:676–82.
9. Kankaanpaa P, Paavolainen L, Tiitta S, Karjalainen M, Paivarinne J, Nieminen J, et al. BioImageXD: an open, general-purpose and high-throughput image-processing platform. Nat Methods 2012;9:683–9.
10. Leinster T, Cobbold CA. Measuring diversity: the importance of species similarity. Ecology 2012;93:477–89.
11. Padfield D, Rittscher J, Roysam B. Spatio-temporal cell segmentation and tracking for automated screening. In: Proceedings of the 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2008 May 14–17; Paris, France. New York, NY: IEEE; 2008.
12. Gough AH, Chen N, Shun TY, Lezon TR, Boltz RC, Reese CE, et al. Identifying and quantifying heterogeneity in high content analysis: application of heterogeneity indices to drug discovery. PLoS One 2014;9:e102678.
13. Aharon M, Elad M, Bruckstein A. k-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Sig Proc 2006;54:4311–22.
14. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. Nature 2013;501:346–54.
15. Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Grainge MJ, Lee AH, et al. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. J Clin Oncol 2011;29:1949–55.
16. Jacob T, Gray JW, Troxell M, Vu TQ. Multiplexed imaging reveals heterogeneity of PI3K/MAPK network signaling in breast lesions of known PIK3CA genotype. Breast Cancer Res Treat 2016;159:575–83.