Letters

RESEARCH LETTER

Optical Biopsy of Bladder Cancer Using Crowd-Sourced Assessment

Crowdsourcing and optical biopsy are emerging technologies with broad applications in clinical medicine and research. Crowdsourcing, an interactive digital platform that uses multiple individual contributions to efficiently perform a complex task, has been successfully used in diverse disciplines ranging from performance assessment in surgery to optimization of tertiary protein conformations.[1,2] Optical biopsy technologies provide real-time tissue imaging with histology-like resolution and the potential to guide intraoperative decision making.[3-5] An example is confocal laser endomicroscopy (CLE), which can be used for the diagnosis and grading of bladder cancer.[6] To further assess the adoptability of optical biopsy as a diagnostic tool, we applied crowdsourcing to determine the barriers to learning how to diagnose cancer using CLE. We hypothesized that a nonmedically trained crowd could learn to rapidly and accurately distinguish between cancer and benign tissue.

Methods | Amazon Mechanical Turk (Amazon.com) users were recruited as the crowd using a software platform developed by C-SATS. Each crowd worker first completed a validated training module[6] and answered a standard screening question, and then assessed a CLE video sequence randomly selected from a set of 12 sequences derived from a benign (n = 3) or cancerous (n = 9) urothelium (**Figure 1**). Videos were previously annotated by an expert user (J.C.L.), and diagnoses were confirmed by pathology under a Stanford University institutional review board–approved protocol. For a video to be categorized as showing a cancerous urothelium, correct classification by at least 70% of the crowd, which is the lowest statistical threshold for differentiation from random guessing, was required. Agreement with the expert user by at least 70% of crowd workers was also used to classify microscopic features with 2 categories (papillary structure, organization, morphology, cellular cohesiveness, and cellular borders). Microscopic vascular features with 3 categories were categorized based on a lower threshold of 35% agreement. Crowd workers were compensated 50¢ for each video assessed and blinded to patient history and diagnosis.

Results | A total of 1283 ratings from 602 crowd workers were received in 9 hours, 27 minutes. A total of 1173 ratings were eligible for analysis based on correct screening response. The crowd accurately distinguished a cancerous urothelium from a benign urothelium in 11 of 12 video sequences (92%) (**Figure 2**). The single erroneous classification was of low-grade bladder cancer. In the assessment of microscopic characteristics, the crowds achieved the highest accuracy for cellular borders (10 of 12 video sequences [83%]), followed by vascularity (9 of 12 video sequences [75%]), organization (8 of 12 video sequences [67%]), and cellular cohesiveness (7 of 12 video sequences [58%]). One video was not included in the analysis of cellular morphology (8 of 11 video sequences [73%]) because it contained both monomorphic and pleomorphic cells, but the crowd workers were not given the option to select both. The diagnostic accuracy was lowest for flat vs papillary characterization (6 of 12 video sequences [50%]).

Discussion | Hurdles for dissemination of new diagnostic technologies in surgery include clinical validation, overcoming the learning curve, and result interpretation. We hypothesized that crowdsourcing may provide an efficient and cost-effective means for technology evaluation and refinement of the educational curriculum. To validate CLE for intraoperative optical biopsy of bladder cancer, we previously found high diagnostic accuracy and moderate interobserver agreement for image interpretation by 15 novice CLE users, including urological surgeons, pathologists, and engineers.[6] Herein, using crowdsourcing, we efficiently expanded our study to a considerably larger crowd. After a brief training module, the crowd achieved an overall diagnostic accuracy of 92% for cancer classification and exceeded 70% accuracy for cellular borders, vasculature, and cellular morphology. The lower accuracy for cellular cohesiveness, organization, and papillary structure suggests a path toward further refinement of the CLE training curriculum. The limitations of our study include a lack of demographic information for crowd workers and a limited number of video sequences. Overall, the diagnostic accuracy achieved with crowdsourcing demonstrates the relative ease of learning an optical imaging technology for enhanced detection of cancer and a complementary strategy to evaluate new surgical technologies.

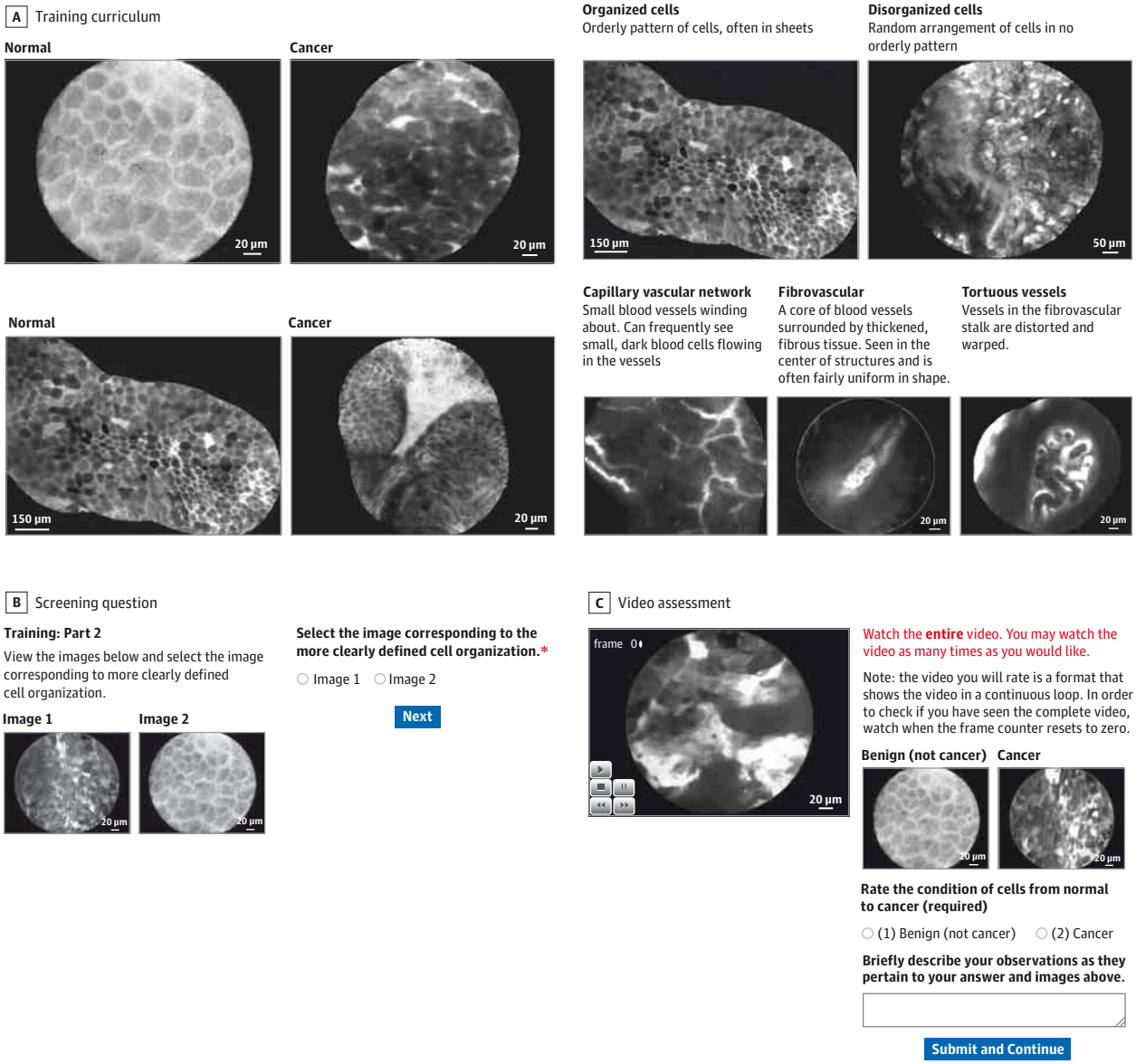
Stephanie P. Chen, BS
Sarah Kirsch, BS
Dimitar V. Zlatev, MD
Timothy Chang, MD
Bryan Comstock, MS
Thomas S. Lendvay, MD
Joseph C. Liao, MD

Author Affiliations: Department of Urology, Stanford University School of Medicine, Stanford, California (Chen, Zlatev, Chang, Liao); Department of Urology, University of Washington, Seattle (Kirsch, Comstock, Lendvay); VA Palo Alto Health Care System, Palo Alto, California (Liao).

Corresponding Author: Joseph C. Liao, MD, Department of Urology, Stanford University School of Medicine, 300 Pasteur Dr, S-287, Stanford, CA 94305-5118 (jliao@stanford.edu).

Published Online: September 30, 2015. doi:10.1001/jamasurg.2015.3121.

Figure 1. Representative Screenshots of Online Modules for Crowd-Sourced Assessment of Bladder Cancer Using Confocal Laser Endomicroscopy (CLE)



A, Each crowd worker was presented with a computer-based CLE training module that included previously validated diagnostic criteria of a cancerous urothelium and a benign urothelium. B, Crowd workers were then asked a test question. An incorrect answer excluded the crowd worker's responses from subsequent analysis. C, Crowd workers were randomly assigned to evaluate 1 of 12 video sequences. The video sequences consisted of 3 benign urothelia and 9 cancerous urothelia (4 low-grade carcinomas and 5 high-grade carcinomas). Crowd workers were asked to designate the video image as cancer or benign, as well as evaluate 6 microscopic features (flat vs papillary, organization, morphology, cellular cohesiveness, cellular borders, and vascularity). Crowd workers could elaborate on their observations with free text responses. Additional CLE videos could be reviewed by reentering the system. Each CLE video received a minimum of 100 responses.

Author Contributions: Ms Chen and Dr Liao had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Chen, Kirsch, Zlatev, Lendvay, Liao.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Chen, Kirsch, Zlatev, Lendvay, Liao.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Zlatev, Comstock.
Administrative, technical, or material support: Chen, Kirsch, Zlatev, Chang, Lendvay, Liao.
Study supervision: Lendvay, Liao.

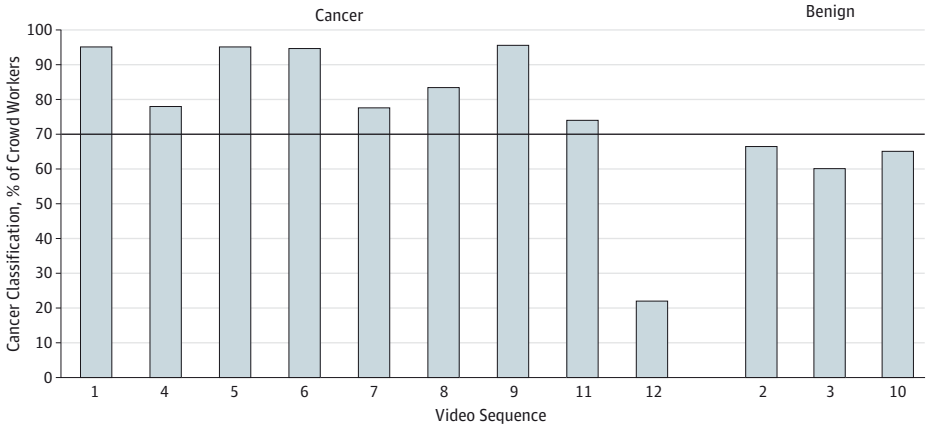
Conflict of Interest Disclosures: Mr Comstock and Dr Lendvay are co-owners of C-SATS.

Funding/Support: Supported in part by Stanford University School of Medicine Medical Scholars Fellowship to Ms Chen.

Role of the Funder/Sponsor: The Stanford University School of Medicine had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We thank Justin Warren, MBA, from C-SATS for his technical support in developing the survey pages for crowd workers.

Figure 2. Diagnostic Accuracy of Crowd-Sourced Assessment of Confocal Laser Endomicroscopy Imaging for Bladder Cancer



To classify a video image as a cancerous urothelium, a threshold of 70% agreement by the crowd was used on the basis that this represented the lowest percentage with a 1-sided 90% CI that excluded a random classification for cancerous vs benign by the crowd. The crowd was able to accurately distinguish between a cancerous and a benign urothelium in 11 of 12 video sequences (92%), with 1 video sequence of (low-grade) cancer incorrectly classified as benign. Diagnostic accuracy was lowest for papillary structure; this provides a presumptive explanation for the single erroneous video classification for which the majority of crowd workers missed the presence of neoplastic papillary features.

1. Ranard BL, Ha YP, Meisel ZF, et al. Crowdsourcing—harnessing the masses to advance health and medicine, a systematic review. *J Gen Intern Med*. 2014;29(1):187-203.

2. Chen C, White L, Kowalewski T, et al. Crowd-Sourced Assessment of Technical Skills: a novel method to evaluate surgical performance. *J Surg Res*. 2014;187(1):65-71.

3. Chen SP, Liao JC. Confocal laser endomicroscopy of bladder and upper tract urothelial carcinoma: a new era of optical diagnosis? *Curr Urol Rep*. 2014;15(9):437.

4. Goetz M, Kiesslich R, Dienes HP, et al. In vivo confocal laser endomicroscopy of the human liver: a novel method for assessing liver microarchitecture in real time. *Endoscopy*. 2008;40(7):554-562.

5. Pogorzelski B, Hanenkamp U, Goetz M, Kiesslich R, Gosepath J. Systematic intraoperative application of confocal endomicroscopy for early detection and resection of squamous cell carcinoma of the head and neck: a preliminary report. *Arch Otolaryngol Head Neck Surg*. 2012;138(4):404-411.

6. Chang TC, Liu JJ, Hsiao ST, et al. Interobserver agreement of confocal laser endomicroscopy for bladder cancer. *J Endourol*. 2013;27(5):598-603.