



Original Article

Crowdsourcing the General Public for Large Scale Molecular Pathology Studies in Cancer



Francisco J. Candido dos Reis^{a,b}, Stuart Lynn^c, H. Raza Ali^d, Diana Eccles^e, Andrew Hanby^f, Elena Provenzano^g, Carlos Caldas^d, William J. Howat^d, Leigh-Anne McDuffus^d, Bin Liu^d, Frances Daley^h, Penny Coulson^h, Rupesh J. Vyasⁱ, Leslie M. Harrisⁱ, Joanna M. Owensⁱ, Amy F.M. Cartonⁱ, Janette P. McQuillanⁱ, Andy M. Patersonⁱ, Zohra Hirjiⁱ, Sarah K. Christieⁱ, Amber R. Holmesⁱ, Marjanka K. Schmidt^j, Montserrat Garcia-Closas^h, Douglas F. Easton^{a,k}, Manjeet K. Bolla^k, Qin Wang^k, Javier Benitez^{l,m}, Roger L. Milne^{n,o}, Arto Mannermaa^p, Fergus Couch^q, Peter Devilee^r, Robert A.E.M. Tollenaar^s, Caroline Seynaeve^w, Angela Cox^t, Simon S. Cross^u, Fiona M. Blows^a, Joyce Sanders^j, Renate de Groot^j, Jonine Figueroa^v, Mark Sherman^v, Maartje Hooning^w, Hermann Brenner^{x,y,z}, Bernd Holczek^{aa}, Christa Stegmaier^{aa}, Chris Lintott^c, Paul D.P. Pharoah^{a,k,*}

^a Department of Oncology, University of Cambridge, Cambridge, UK

^b Department of Gynecology and Obstetrics, Ribeirao Preto School of Medicine, University of Sao Paulo, Ribeirao Preto, Brazil

^c Department of Physics (Astrophysics), University of Oxford, Oxford, UK

^d Cancer Research UK, Cambridge Institute, Cambridge, UK

^e University of Southampton, Southampton, UK

^f University of Leeds, Leeds, UK

^g Addenbrookes Hospital NHS Trust, Cambridge, UK

^h Institute of Cancer Research, London, UK

ⁱ Cancer Research UK, London, UK

^j Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

^k Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

^l Human Genotyping (CEGEN) Unit, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

^m Biomedical Network on Rare Diseases (CIBERER), Madrid, Spain

ⁿ Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia

^o Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Australia

^p Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Finland

^q Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

^r Department of Human Genetics & Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands

^s Department of Surgical Oncology, Leiden University Medical Center, Leiden, The Netherlands

^t Sheffield Cancer Research, Department of Oncology, University of Sheffield, Sheffield, UK

^u Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK

^v National Cancer Institute, USA

^w Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

^x Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany

^y Division of Preventive Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany

^z German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

^{aa} Saarland Cancer Registry, Saarbrücken, Germany

ARTICLE INFO

Article history:

Received 27 March 2015

Received in revised form 7 May 2015

Accepted 8 May 2015

Available online 9 May 2015

Keywords:

Citizen science

Crowd science

Crowdsourcing

Breast cancer

ABSTRACT

Background: Citizen science, scientific research conducted by non-specialists, has the potential to facilitate biomedical research using available large-scale data, however validating the results is challenging. The Cell Slider is a citizen science project that intends to share images from tumors with the general public, enabling them to score tumor markers independently through an internet-based interface.

Methods: From October 2012 to June 2014, 98,293 Citizen Scientists accessed the Cell Slider web page and scored 180,172 sub-images derived from images of 12,326 tissue microarray cores labeled for estrogen receptor (ER). We evaluated the accuracy of Citizen Scientist's ER classification, and the association between ER status and prognosis by comparing their test performance against trained pathologists.

Findings: The area under ROC curve was 0.95 (95% CI 0.94 to 0.96) for cancer cell identification and 0.97 (95% CI 0.96 to 0.97) for ER status. ER positive tumors scored by Citizen Scientists were associated with survival in a

* Corresponding author at: Departments of Oncology and Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worth's Causeway, Cambridge CB1 8RN, England, UK.

E-mail address: pp10001@medschl.cam.ac.uk (P.D.P. Pharoah).

similar way to that scored by trained pathologists. Survival probability at 15 years were 0.78 (95% CI 0.76 to 0.80) for ER-positive and 0.72 (95% CI 0.68 to 0.77) for ER-negative tumors based on Citizen Scientists classification. Based on pathologist classification, survival probability was 0.79 (95% CI 0.77 to 0.81) for ER-positive and 0.71 (95% CI 0.67 to 0.74) for ER-negative tumors. The hazard ratio for death was 0.26 (95% CI 0.18 to 0.37) at diagnosis and became greater than one after 6.5 years of follow-up for ER scored by Citizen Scientists, and 0.24 (95% CI 0.18 to 0.33) at diagnosis increasing thereafter to one after 6.7 (95% CI 4.1 to 10.9) years of follow-up for ER scored by pathologists.

Interpretation: Crowdsourcing of the general public to classify cancer pathology data for research is viable, engages the public and provides accurate ER data. Crowdsourced classification of research data may offer a valid solution to problems of throughput requiring human input.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The assessment of tissue protein expression by immunohistochemistry (IHC) is widely used in both the clinical and the research settings. IHC combined with tissue microarray (TMA) technology (Wan et al., 1987; Kononen et al., 1998) provides an efficient approach to the study of multiple molecular markers in hundreds or thousands of tumors. TMAs are produced by removing cylindrical cores of tissue from up to donor paraffin blocks and embedding these into a single recipient paraffin block at set array coordinates. Several hundreds of tumors may be embedded in a single TMA. This has the potential to reduce inter-assay variability and to reduce the cost of research (Camp et al., 2008). Consequently, the large sample sizes required for robust inference in clinical epidemiology are achievable. A typical study may include over 10,000 cases (Ali et al., 2014). However, the process still relies on manual scoring of labeled sections by trained researchers. This is time consuming and scoring remains a rate-limiting step in this type of research. One solution to this bottleneck is to scan the labeled sections and to use automated analysis of the digitized images of each core. Several image analysis algorithms have been shown to perform reasonably well for some IHC markers (Giltane and Rimm, 2004; Bolton et al., 2010; Ali et al., 2013; Howat et al., 2015). While automated image analysis remains promising, its implementation may be complex and it has not yet replaced manual scoring in large scale molecular epidemiology studies in cancer.

An alternative approach to automated image analysis is crowdsourcing in which a function – here scoring of IHC labeled sections of tumor cores – is outsourced to an undefined and generally large group of people in the form of an open call. The crucial prerequisites are the use of the open call format and the large network of potential contributors (Howe, 2006). Crowdsourcing relies on parallel independent inputs from individuals allowing for large group size, maximizing cognitive diversity and enhancing group performance (Page, 2008).

The Citizen Science Alliance (<http://www.citizensciencealliance.org>) is a collaboration of scientists, software developers and educators, who use the concept of crowdsourcing to develop, manage and utilize internet-based citizen science projects in order to further scientific research and to promote the public understanding of science. Through citizen science projects, thousands of Citizen Scientists have collected, organized and classified data for research purposes. Some successful initiatives are: the investigation of galaxy morphology (Lintott et al., 2008), the prediction of protein structures (Cooper et al., 2010) and the alignment of multiple sequences in genomic studies (Kawrykow et al., 2012). The Cell Slider project was established to enable the scoring of tumors labeled using IHC by untrained members of the general public – Citizen Scientists – through an internet-based interface. In this paper we report the results of the first Cell Slider project in which Citizen Scientists scored estrogen receptor (ER) expression in images of tumor cores from a large number of breast cancers arrayed in TMAs.

2. Methods

2.1. Study Design, Setting, and Population

This study was performed using pathology data from the Breast Cancer Association Consortium (BCAC), an international collaboration that was established to provide large sample sizes for examining risk factors, genetic associations and prognostic markers in breast cancer (Breast Cancer Association Consortium, 2006). The BCAC resource comprised 12,326 scanned images from breast cancer TMA cores stained for estrogen receptor (ER). A total of 3082 cores from the SEARCH study (Lesueur et al., 2005) – that had been previously scored by the same pathologist under conventional microscopy and without access to patient clinical records. The cores were from the tumors of 6378 patients from 10 studies (Appendix 1). Information on clinic-pathological characteristics of each patient was obtained from clinical records or centralized review of case notes. This included ER status for which was either taken from independent research-based pathology review or, where this had not been carried out, from the clinical records. Relevant research ethics committees approved all the studies and samples were anonymized before being sent to two coordinating centers at Strangeways Research Laboratory (University of Cambridge, UK) and the Breakthrough Pathology Core Facility (Institute of Cancer Research, London, UK) for analysis. Fig. 1 summarizes the study design.

TMA sections were immunostained in several centers and each stained TMA slide was digitized using the Ariol platform (Genetix Ltd, Hampshire, UK) and high-resolution images of each tumor core were subsequently extracted for analysis. The ease of scoring of TMA images before and after transforming the colors in a variety of combinations was evaluated by beta-testing by experienced Citizen Science Alliance volunteers. The preferred colors were then used subsequently. The colors of the images were transformed using the ImageMagick library. The colors of the image were first negated (replacing each pixel by its complementary color) and then the saturate was increased by 300% and the hue reduced by 82%. Finally the full image was divided into 16 sub-images which were resized to 495 by 496 pixels each. The four corner sub-images were removed as they often had no tumor material present and the remaining 12 sub-images were uploaded to the Cell Slider project web site.

2.2. Citizen Scientist Training and Scoring

Any member of the public (Citizen Scientists) can participate of the project at <http://www.cellslider.net/>. Once in the website the Citizen Scientist can register a user name and a password or proceed without registration. At first entry the Citizen Scientist is provided with a brief web based training tutorial in which the task and key steps required to score each image are described. After completing the training the Citizen Scientist is presented with an image to

Download English Version:

<https://daneshyari.com/en/article/2121176>

Download Persian Version:

<https://daneshyari.com/article/2121176>

[Daneshyari.com](https://daneshyari.com)