Research Spending & Results

Award Detail

Awardee: BROAD INSTITUTE, INC., THE
Doing Business As Name: Broad Institute, Inc.
PD/PI: Anne Carpenter
(617) 714-7130
anne@broadinstitute.org
Co-PD(s)/co-PI(s): Paul Rees
Award Date: 07/16/2015
Estimated Total Award Amount: $ 394,349
Funds Obligated to Date: $ 394,349
Start Date: 07/15/2015
End Date: 06/30/2018
Transaction Type: Grant
Agency: NSF
Awarding Agency Code: 4900
Funding Agency Code: 4900
CFDA Number: 47.074
Primary Program Source: 040100 NSF RESEARCH & RELATED ACTIVIT
Award Title or Description: Bilateral BBSRC-NSF/BIO: Mining of imaging flow cytometry data for label-free, single cell analysis
Federal Award ID Number: 1458626
DUNS ID: 623544785
Parent DUNS ID: 623544785
Program: ADVANCES IN BIO INFORMATICS
Program Officer: Peter McCartney
(703) 292-8470
pmccartn@nsf.gov

Awardee Location

Street: 415 Main Street
City: Cambridge
State: MA
ZIP: 02142-1401
County: Cambridge
Country: US
Awardee Cong. District: 07

Primary Place of Performance

Organization Name: Broad Institute, Inc.
Street: 7 Cambridge Center
City: Cambridge
State: MA
ZIP: 02142-1401
County: Cambridge
Country: US
Abstract at Time of Award

The goal of this project is to develop and demonstrate software to mine data from imaging flow cytometers. These instruments can capture thousands of images of cells per second. The images can in theory be analyzed to precisely measure hundreds of features related to a cell's appearance ("morphology"); this project is to develop advanced machine-learning software to accomplish this, unlocking the otherwise hidden information within the images. The software will be developed, improved, and validated in several demonstration experiments involving the cell cycle, the component cells of primary blood, immune cell activation, and stem cell identity. The goal will be to use as few or indeed no fluorescent biomarkers, eliminating the need to perturb cells. The resulting open-source software will be freely available to scientists worldwide for both applied and clinical research, and will be accompanied by user-friendly training materials and in-person workshops. The project is collaborative and interdisciplinary and includes training early career-stage scientists in computational biology, via the existing Scientists without Borders program.

The project involves close collaboration with researchers using imaging flow cytometers and builds on successful interdisciplinary work in biological data mining. In order to devise the novel software and methodology to mine the large datasets acquired using imaging flow cytometry, the team will develop algorithms to seamlessly import data from an imaging cytometer, robustly segment cells, quality-filter them (e.g., for debris and blur), and quantify morphological parameters (usually hundreds) for each cell (usually thousands), including various measures of size, shape, and texture. Using these features, trained machine-learning algorithms will identify cell phenotypes of interest or otherwise characterize the state of cell in driving biological projects from project partners who use imaging flow cytometry in a host of biological research studies. The goal will be to use as few or indeed no fluorescent biomarkers, eliminating the need to perturb cells. The project will give the scientific community a validated, open-source software toolbox of image-processing and machine learning algorithms readily usable by biologists. More information about the project can be found at: http://www.broadinstitute.org/~anne/Carpenter_NSF_ImagingFlowCytometry.html

Publications Produced as a Result of this Research

Note: When clicking on a Digital Object Identifier (DOI) number, you will be taken to an external site maintained by the publisher. Some full text articles may not yet be available without a charge during the embargo (administrative interval).

Some links on this page may take you to non-federal websites. Their policies may differ from this site.

Thomas Blasi, Holger Hennig, Huw D. Summers, Fabian J. Theis, Joana Cerveira, James O. Patterson, Derek Davies, Andrew Filby, Anne E. Carpenter & Paul Rees "Label-free cell cycle analysis for high-throughput imaging flow cytometry" NATURE COMMUNICATIONS, v.7, 2016, p.1. doi:10.1038/ncomms10256

Project Outcomes Report

Disclaimer

This Project Outcomes Report for the General Public is displayed verbatim as submitted by the Principal Investigator (PI) for this award. Any opinions, findings, and conclusions or recommendations expressed in this Report are those of the PI and do not necessarily reflect the views of the National Science Foundation; NSF has not approved or endorsed its content.

Microscopy is a fundamental technology routinely used by researchers to examine subtle details of cell structures to answer biological questions. However, it is rather slow: images of millions of cells cannot be digitally captured efficiently and it is challenging to examine cells that overlap each other or do not attach to a microscope slide. For cells in suspension, such as in blood samples, flow cytometry is more suitable. A conventional flow cytometer allows scientists to collect information on thousands of cells per second. However, it captures only a single intensity value for each fluorescent marker used. A newer generation of flow cytometers equipped with sensitive cameras can now capture entire images of cells, hence the name imaging flow cytometry. Imaging flow cytometry beats its non-imaging predecessor in terms of resolution while still measuring thousands cell per second.

We hypothesized that imaging flow cytometry would be most useful if it could be analyzed with the latest computational algorithms, such as machine learning. In fact, we even wondered if we could eliminate the need for fluorescent labeling of cells entirely by extracting cell measurements straight from images taken with plain light. If we could do this, we could preserve the true states of the cells more faithfully, while eliminating the tedious and expensive steps of identifying the appropriate marker, and performing a sophisticated laboratory procedure involving multiple rounds of sample preparations (such as fixing the cells, antibody incubations, washing, centrifugation and resuspension).

We therefore aimed to use machine learning to capture the hidden information within the images, in order to eliminate or reduce the need for fluorescent biomarkers that can perturb cells. To this end, we built a computational workflow called Deepometry that includes all major steps of the analysis, integrating state-of-the-art computer vision methodologies, with a special focus on deep learning, a powerful type of machine learning. Our workflow is flexible for many biologists’ experiments, providing solutions for many kinds of assays and experiments, including working pipelines for an unlimited number of color channels and various object sizes, data structures, and image formats.

We optimized and proved the usefulness of the approach in real-world biological and clinical applications. These include cell cycle analysis, hematological disease, quantification of red blood cell morphology, and immune system activation, including
allergic eosinophil/neutrophil activation. The software and educational materials we produced in this project are freely available so they can be used, and further advanced, by scientists around the world.