search by choosing images from the results, marking them as positive (meaning 'retrieve more images similar to these') and negative ('exclude images similar to these') and repeating until satisfied. A stand-alone client that does not require a local copy of OMERO is also available (**Supplementary Software**). It permits users to choose images on their local computer, calculate features and find similar images in remote databases that have OMERO.searcher installed (**Supplementary Note**). (The next release of OMERO.searcher will support searching across multiple OMERO databases at different locations, assuming access rights.)

To test how well the searcher retrieves relevant images, we performed tests using two distinct fluorescence microscopy databases, PSLID and The Cell: An Image Library (The Cell; http://www.cellimagelibrary.org/). We created classes of images with the same content annotations and ranked the images by similarity to one or more query images drawn from one of those classes (**Supplementary Methods**). We measured success using the area under a receiver operating characteristic curve, in which retrieval rates for images from the desired class are compared to those for images from undesired classes. We obtained good results for many different patterns from both databases (**Fig. 1**) even though The Cell contained images captured at different resolutions and from different microscope types. Increasing the number of images in the query improved result quality, as did using both positive and negative examples for the same total number of labeled images (**Fig. 1b**). The images used in this second test were collected at 40× magnification. We obtained similar results when searching with downsampled versions to simulate a query with images collected at 10× magnification (**Supplementary Fig. 2**). Feature sets are also available to permit searching with three-dimensional images and time series.

OMERO.searcher is an open-source content-based image search tool for the cell and computational biology community. It has several useful applications, such as asking whether someone has previously observed a pattern similar to an unrecognized one or for finding examples of a particular pattern in other cell types or different modes of microscopy.

Note: Supplementary information is available at http://www.nature.com/doifinder/10.1038/nmeth.2086.

ACKNOWLEDGMENTS

This research was supported in part by US National Institutes of Health grants GM075205, EB008516 and GM092708 and by grant 095931 from the Wellcome Trust. B.H.C. was supported by a postdoctoral fellowship from the Korea Research Foundation Grant (KRF-2008-D00316). We thank K. Eliceiri, J. Swedlow, J. Moore, D. Orloff, L. Wu and C. Faloutsos for helpful discussions.

AUTHOR CONTRIBUTIONS

B.H.C. and J.A.B. performed research and contributed code, R.F.M. conceived and guided research, I.C.-B. contributed code, and B.H.C. and R.F.M. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Baek Hwan Cho[1], Ivan Cao-Berg[1], Jennifer Ann Bakal[2] & Robert F Murphy[1–5]

[1]Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. [2]Center for Bioimage Informatics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. [3]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. [4]Department of Machine Learning, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. [5]Freiburg Institute for Advanced Studies, Albert Ludwig University of Freiburg, Germany. e-mail: murphy@cmu.edu

1. Swedlow, J.R. *Nat. Cell Biol.* **13**, 183 (2011).
2. Faloutsos, C. *et al. J. Intell. Inf. Syst.* **3**, 231–262 (1994).
3. Allan, C. *et al. Nat. Methods* **9**, 245–253 (2012).
4. Glory, E. & Murphy, R.F. *Dev. Cell* **12**, 7–16 (2007).
5. Wu, L., Faloutsos, C., Sycara, K.P. & Payne, T.R. in *Proc. 26th Int. Conf. Very Large Data Bases* (eds., Abbadi, A.E. *et al.*) 297–306 (Morgan Kaufmann, 2000).
6. Huang, K., Lin, J., Gajnak, J.A. & Murphy, R.F. in *Proc. 2002 IEEE Int. Symp. Biomed. Imaging*, 325–328 (2002).

SimuCell: a flexible framework for creating synthetic microscopy images

To the Editor: Advances in high-content fluorescence microscopy have driven the development of analytical approaches for extracting meaningful information from rich and complex biological image data. Algorithm development can be aided dramatically by using curated test data. To evaluate the generality and performance of new algorithms, test data should contain annotation on how images differ in terms of cell phenotypes, population heterogeneity and/or microenvironmental[1] effects. Currently there is a paucity of diverse, well-annotated data. A complementary approach is to use synthetically generated data, in which biological[1] and imaging[2] effects can be varied independently and 'ground truths' are known. Although approaches exist for rendering realistic cells[3,4], creating biologically realistic cell-population images has remained challenging: biomarker, cell and population phenotypes can be subtle, interconnected and system dependent. To deal with these challenges, we developed SimuCell (**Supplementary Software**; updated software available at http://www.SimuCell.org/), an open-source framework (**Fig. 1a**) for specifying and rendering realistic microscopy images containing diverse cell phenotypes, heterogeneous populations, microenvironmental dependencies and imaging artifacts.

SimuCell differs from existing cell-population generators[5] in three ways. First, SimuCell can generate heterogeneous cellular populations composed of diverse cell types. Each cell type can be defined independently by specifying models for cell and organelle shape and distributions of markers over these shapes. Models are typically algorithmic, but there is support for rendering produced by other tools, such as the highly realistic models learned from image data by CellOrganizer[3] (via the new SLML markup language). Second, SimuCell allows users to specify interdependencies among population, biomarker and cell phenotypes. For example, a marker's cellular distribution can be affected by the cell's microenvironment (**Fig. 1b**, marker 1) as well as the localization pattern of another marker (**Fig. 1b**, markers 2 and 3). These definable image properties are accessible to users either via a novel scripting syntax built on top of MATLAB or through a graphical user interface; intermediate results can be used to define further 'ground truths' (for example, cell boundaries can be used to validate segmentation algorithms). Finally, SimuCell is easily extensible, providing a standard framework for defining new plugins that can also be shared through the SimuCell website. Users interested in adding novel phenotypes to SimuCell's palette can typically do so by writing just a few lines of code, in part because of MATLAB's extensive library of functions. We also intend to implement a user forum to share ideas, scripts, plugins and images. Thus SimuCell allows the definition of a broad

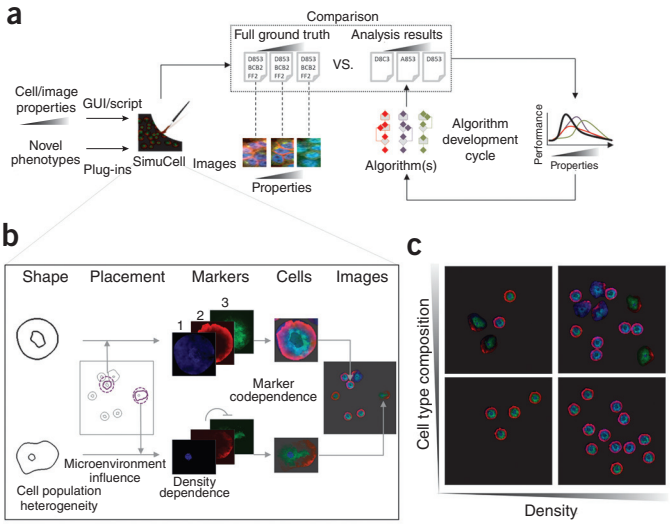

Figure 1 | Synthetic image generation using SimuCell. (**a**) Typical workflow of SimuCell use during algorithm development. (**b**) Steps involved in generation of synthetic images. In this example, microenvironment (local cell density) affects marker 1, and marker 2 influences marker 3. (**c**) SimuCell can be used to create images in which cell-population properties are varied independently.

range of phenotypes, encompassing nontrivial population-level effects such as cell-type heterogeneity or local cell-density effects (**Fig. 1c**). Although realistic synthetic data cannot replace true experimental data[6], SimuCell can be a useful part of the algorithm developer's toolbox by generating rich, flexible test image data sets containing specified, parameterized 'biological' effects.

Note: Supplementary information is available at http://www.nature.com/doifinder/10.1038/nmeth.2096.

ACKNOWLEDGMENTS

We thank R. Murphy and members of the Altschuler and Wu laboratories for helpful feedback and discussions. This research was supported by US National Institutes of Health R01 grants (GM085442 to S.J.A. and GM081549 to L.F.W.), the Welch Foundation (I-1619 to S.J.A. and I-1644 to L.F.W.), CPRIT RP10900 (L.F.W.) and the University of Texas Southwestern Medical Center (QP-SURF to N.E.F.H.).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Satwik Rajaram[1,2,4], Benjamin Pavie[1,2,4], Nicholas E F Hac[3], Steven J Altschuler[1,2] & Lani F Wu[1,2]

[1]Green Center for Systems Biology, University of Texas Southwestern Medical Center, Dallas, Texas, USA. [2]Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas, USA. [3]University of Virginia School of Medicine, Charlottesville, Virginia, USA. [4]These authors contributed equally to the work.
e-mail: steven.altschuler@utsouthwestern.edu or lani.wu@utsouthwestern.edu

1. Snijder, B. *et al. Nature* **461**, 520–523 (2009).
2. Bray, M.A., Fraser, A.N., Hasaka, T.P. & Carpenter, A.E. *J. Biomol. Screen.* **17**, 266–274 (2012).
3. Zhao, T. & Murphy, R.F. *Cytometry A* **71A**, 978–990 (2007).
4. Svoboda, D., Kozubek, M. & Stejskal, S. *Cytometry A* **75A**, 494–509 (2009).
5. Lehmussola, A., Ruusuvuori, P., Selinummi, J., Huttunen, H. & Yli-Harja, O. *IEEE Trans. Med. Imaging* **26**, 1010–1016 (2007).
6. Anonymous. *Nat. Methods* **8**, 885 (2011).

PhenoRipper: software for rapidly profiling microscopy images

To the Editor: Recent advances in fluorescence microscopy have enabled unprecedented progress in many areas of biology. With the technology to perform high-content image-based screens now accessible to many labs, the analysis of the resulting large and complex data sets has become a bottleneck. Existing image analysis platforms[1–3] offer flexible and sophisticated toolboxes for extracting biological information from image data. However, they can require steep learning curves, tuning of many parameters and long computational runtimes. There is an unmet need for easy-to-use tools that enable bench scientists to rapidly interpret their image data sets. Here we describe PhenoRipper (**Supplementary Software**; updated versions available at http://www.phenoripper.org/), an open-source software tool designed for rapid exploration of high-content microscopy images (**Fig. 1a** and **Supplementary Fig. 1**). PhenoRipper permits rapid and intuitive comparison of images obtained under different experimental conditions based on image phenotype similarity.

To minimize user input, PhenoRipper automatically identifies features from the images; users may only be required to modify default values of a few, visually interpretable, parameters. To increase speed, we chose a segmentation-free approach[4,5]: the software breaks images down into a square grid of blocks[6–8] and performs analysis on these blocks rather than on individual cells. To capture heterogeneity, PhenoRipper identifies characteristic patterns of neighboring blocks and describes each image in terms of the occurrence frequencies of these patterns[6,8]. Finally, a simple graphical user interface, PhenoBrowser, is used to tie together images, features and profiles. Profiles can be annotated or combined (for example, by experimental or replicate conditions) to help interpret and explore their visual grouping. These design choices let users analyze their images an order of magnitude faster than existing unsupervised platforms (**Supplementary Fig. 2**). PhenoRipper does not replace traditional single cell–based analysis approaches[2,9,10] as it does not quantify properties such as area or average nuclear biomarker intensity. Nevertheless, the statistical properties of subcellular-scale phenotypes captured by PhenoRipper can be sufficient to accurately group cellular perturbations and identify outliers (**Supplementary Fig. 3a**).

PhenoRipper's engine performs four major steps (**Fig. 1a** and **Supplementary Fig. 1**). (i) PhenoRipper identifies foreground blocks. Images are gridded to a user-specified block size (20–30 blocks per cell works well), and blocks are selected when the intensities of >50% of their pixels exceed a foreground threshold. This threshold is precalculated based on a small subset of images (**Supplementary Methods**), but it can easily be changed by the user. (ii) PhenoRipper identifies the most common foreground block types. To do this, it characterizes blocks by their distributions of assigned pixel colors and applies cluster analysis to classify them into different block types. This measurement is not sensitive to cell orientation and captures more information than simple averages (for example, a block with 50% red and 50% blue pixels would be different from a block with 100% purple pixels). (iii) PhenoRipper uses cluster analysis to identify superblock types, which represent the most common block type co-occurrence patterns within 3×3 block neighborhoods. The use of blocks and superblocks helps