

Published in final edited form as:

*Nat Methods*. ; 9(7): . doi:10.1038/nmeth.2096.

## SimuCell: a flexible framework for creating synthetic microscopy images

Satwik Rajaram<sup>1,†</sup>, Benjamin Pavie<sup>1,†</sup>, Nicholas Hac<sup>2</sup>, Steven J. Altschuler<sup>1,\*</sup>, and Lani F. Wu<sup>1,\*</sup>

<sup>1</sup>Green Center for Systems Biology and Department of Pharmacology, University of Texas Southwestern Medical Center, 6001 Forest Park Road, ND 9.214, Dallas, Texas 75390, USA.

<sup>2</sup>University of Virginia School of Medicine, P.O. Box 800793, Charlottesville, VA 22908, USA.

Advancements in high-content fluorescence microscopy have driven development of analytical approaches for extracting meaningful information from rich and complex biological image data. Algorithm development can be aided dramatically by the use of curated test data. To evaluate the generality and performance of new algorithms, test data should ideally contain annotation for how images differ in terms of cell phenotypes, population heterogeneity, and/or micro-environmental<sup>1</sup> effects. Currently there is a paucity of diverse, well-annotated data. A complementary approach is to make use of synthetically generated data, in which biological<sup>1</sup> and imaging<sup>2</sup> effects can be varied independently and “ground truths” known. While approaches exist for rendering realistic cells<sup>3,4</sup>, 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 (<http://www.SimuCell.org>), an open-source framework (Fig. 1a) for specifying and rendering realistic microscopy images containing diverse cell phenotypes, heterogeneous populations, micro-environmental dependencies and imaging artifacts.

SimuCell differs from existing cell population generators<sup>5</sup> in three ways. Firstly, 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<sup>3</sup> (via the new SLML markup language). Secondly, SimuCell allows users to specify interdependencies between 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, while intermediate results can define further “ground truths” (e.g. cell boundaries can be used to validate segmentation algorithms).

Finally, SimuCell was designed to be 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 due to MATLAB’s extensive library of functions. Taken together, SimuCell allows the definition of a broad range of phenotypes, encompassing highly non-

\*Corresponding authors = [steven.altschuler@utsouthwestern.edu](mailto:steven.altschuler@utsouthwestern.edu), [lanf.wu@utsouthwestern.edu](mailto:lanf.wu@utsouthwestern.edu).

†These authors contributed equally to the work.

trivial population-level effects such as cell-type heterogeneity or local cell-density effects (Fig. 1c). While realistic synthetic data cannot replace true experimental data<sup>6</sup>, 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.

## Acknowledgments

We thank R. Murphy and members of the Altschuler and Wu labs for helpful feedback and discussions. This research was supported by the National Institute of Health grants R01s (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.), and UTSW QP-SURF to N.H.

## References

1. Snijder B, et al. *Nature*. 2009; 461(7263):520–523. [PubMed: 19710653]
2. Bray MA, Fraser AN, Hasaka TP, Carpenter AE. *Journal of Biomolecular Screening*. 2012; 17(2): 266–274. [PubMed: 21956170]
3. Zhao T, Murphy RF. *Cytometry Part A*. 2007; 71A(12):978–990.
4. Svoboda D, Kozubek M, Stejskal S. *Cytometry Part A*. 2009; 75A(6):494–509.
5. Lehmussola A, Ruusuvoori P, Selinummi J, Huttunen H, Yli-Harja O. *Ieee Transactions on Medical Imaging*. 2007; 26(7):1010–1016. [PubMed: 17649914]
6. *Nat Methods*. 2011; 8(11):885. [PubMed: 22148151]

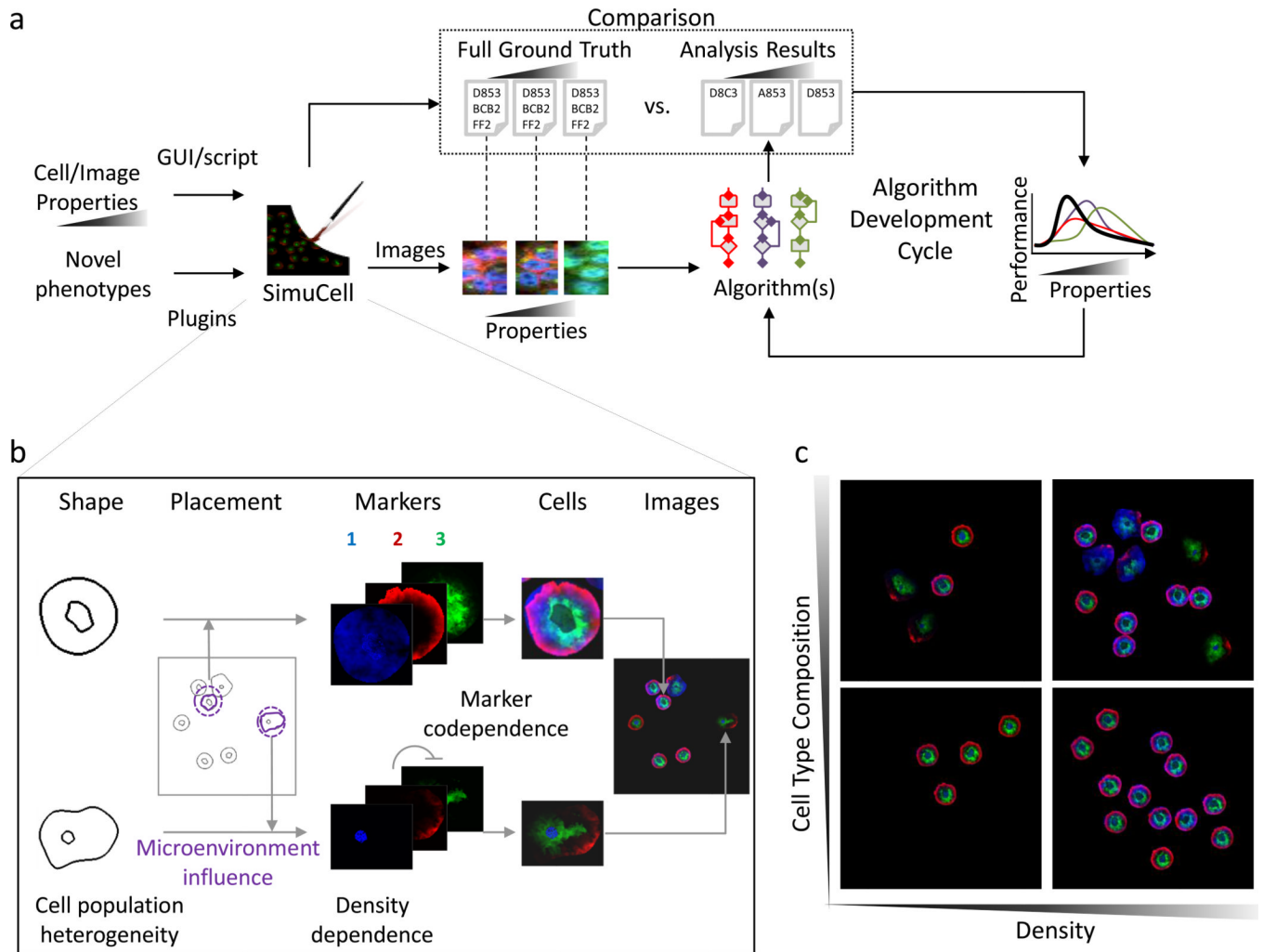


Figure 1.