Published in final edited form as:

Nat Methods. ; 9(7): . doi:10.1038/nmeth.2097.

PhenoRipper: software for rapidly profiling microscopy images

Satwik Rajaram^{1,2}, Benjamin Pavie^{1,2}, Lani F. Wu^{1,2,*}, and Steven J. Altschuler^{1,2,*}

¹Green Center for Systems Biology, University of Texas Southwestern Medical Center, 6001 Forest Park Road, ND 9.214, Dallas, Texas 75390, USA

²Department of Pharmacology, University of Texas Southwestern Medical Center, 6001 Forest Park Road, ND 9.214, Dallas, Texas 75390, USA

Recent advances in fluorescence microscopy have enabled unprecedented progress in many areas of biology. With technology to perform high-content image-based screens now accessible to many labs, analysis of the resulting large and complex data sets has become a bottleneck. Existing image analysis platforms^{1–3} provide 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 datasets. Here we describe PhenoRipper (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 similarity of image phenotypes.

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 gain speed, we chose a segmentation-free approach^{4,5}; images are broken into a square grid of blocks^{6–8} and subsequent analysis is performed on these blocks rather than on individual cells. To capture heterogeneity, characteristic patterns of neighboring blocks are identified, and each image is described in terms of the occurrence frequencies of these patterns^{6,8}. Finally, a simple graphical user interface, PhenoBrowser, ties 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 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 as well as identify outliers (which may be genuine biological “hits” or objects of scrutiny for quality control; Supplementary Fig. 3a).

PhenoRipper’s engine performs four major steps (Fig. 1a and Supplementary Fig. 1). 1) 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 pre-calculated on a small subset of images (Supplementary Methods), but can easily be changed by the user. 2) PhenoRipper identifies the most common foreground block types. Blocks are characterized by their fractions of pixel colors. Cluster analysis is then applied to the foreground blocks to classify them into different block types. This measurement is not sensitive to cell orientation and

*Corresponding authors: lan.wu@utsouthwestern.edu, steven.altshuler@utsouthwestern.edu.

captures more information than simple averages (for example, a block with 50% red and 50% blue pixels would be different than a block with 100% purple pixels). 3) PhenoRipper uses cluster analysis to identify superblock types, representing the most common block type co-occurrence patterns within 3×3 block neighborhoods. The use of blocks and superblocks helps to capture information over different distance scales. To speed up the steps described above, this initial analysis randomly samples a subset of images (Supplementary Methods). 4) PhenoRipper profiles each image by the frequency of occurrence of superblock types. Profiles of experimental conditions are computed by averaging the superblock fractions of their corresponding images. We have found that similarities between profiles are relatively insensitive to parameter variation; Supplementary Figs. 3b and 4. These profiles provide compact, human- and machine-interpretable summaries of image phenotypes. Profile similarities can be used to infer relationships among experimental conditions and underlying mechanisms of perturbations.

We tested PhenoRipper on a data set (640 4-channel images) in which cells were hard to segment and phenotypically heterogeneous⁹ (Fig. 1b, Supplementary Fig. 5). This dataset consists of images of 3T3-L1 preadipocytes monitored for multiple readouts of adipogenesis at different days post induction of differentiation. Our original study, where image analysis was carried out by traditional single-cell analysis, required a tedious manual step of discarding poorly segmented cells. By contrast, PhenoRipper completed its analysis in ~6.5 minutes, selecting image features that could distinguish images from different days of differentiation and identifying superblock types that corresponded roughly to subcellular features of previously identified subpopulations, representing stages of the differentiation process.⁸ Thus, PhenoRipper can reveal meaningful features from heterogeneous populations and images for which robust cell segmentation is not easily achieved.

Next, we reanalyzed a dataset (~10⁵ 3-channel images) whose scale and complexity is representative of high-throughput screens, which typically require dedicated image analysis platforms and analysis expertise¹⁰ (Fig 1c). This dataset is from an experiment in which the effects of ~23,000 genome-wide RNAi-mediated knockdowns on HeLa cells were monitored using cytoskeletal markers. The previous analysis was reported to take over 300 CPU hours, which excludes time required to optimize this analysis pathway. In comparison, PhenoRipper completed analysis of this dataset in ~13 hours on a test desktop, without the need to tune any parameters other than threshold intensity and block size. To compare the profiling results, we focused on the “hits” reported in the previous study (analysis of these ~7000 images took ~30 minutes). Visual grouping of PhenoRipper profiles, annotated by phenotypic classes defined in the previous study, suggested that similarities between knockdown profiles had been largely preserved between the two methods (Fig 1c). Overall, similar profile pairs from PhenoRipper showed strong enrichment for similar biological function (Supplementary Methods and Supplementary Fig. 6). Thus, PhenoRipper provides an approach for rapidly extracting biologically meaningful information from large, complex datasets.

PhenoRipper is designed to serve as an unsupervised, exploratory tool for analysis of fluorescence microscopy images for both novices and experts. It may not always be the optimal tool—some applications may require quantification of specific features on single cells or may be more suitable for supervised classification. Nevertheless, the speed and simplicity of PhenoRipper should make it a highly useful tool for bench-scientists to perform rapid analysis of image data soon after acquisition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank R. Murphy, M. Slack, R. J. Steininger III, C. A. Thorne, P. Xie, and members of the Altschuler and Wu labs for helpful feedback and discussions. This research was supported by the National Institute of Health grants R01 GM085442 (S.J.A.), R01 GM081549 (L.F.W.), and the Welch Foundation I-1619 (S.J.A.) and I-1644 (L.F.W.).

References

1. Collins TJ. *Biotechniques*. 2007; 43(1 Suppl):25–30. [PubMed: 17936939]
2. Carpenter AE, et al. *Genome Biol*. 2006; 7(10):R100. [PubMed: 17076895]
3. Shamir L, Delaney JD, Orlov N, Eckley DM, Goldberg IG. *Plos Computational Biology*. 2010; 6(11)
4. Huang, K.; Murphy, RF. 2004 2nd Ieee International Symposium on Biomedical Imaging: Macro to Nano; 2004. p. 1139-1142.
5. Shamir L, et al. *Source Code Biol Med*. 2008; 3:13. [PubMed: 18611266]
6. Bhattacharya, A., et al. Fifth IEEE International Conference on Data Mining, Proceedings; 2005. p. 50-57.
7. Nanni L, Lumini A. *Artificial Intelligence in Medicine*. 2008; 43(2):87–97. [PubMed: 18440791]
8. Cruz-Roa, A.; Caicedo, JC.; González, FA. *Artificial Intelligence in Medicine*. 2011.
9. Loo LH, et al. *J Cell Biol*. 2009; 187(3):375–384. [PubMed: 19948481]
10. Fuchs F, et al. *Mol Syst Biol*. 2010; 6:370. [PubMed: 20531400]

Figure 1.

(a) Flow chart of analysis performed by PhenoRipper. (b) PhenoBrowser interface. Upper left: 3-D MDS plot of profiles for images of 3T3-L1 cells on different days of differentiation to adipocytes (days indicated by color). Right column: two selected images from day 15 (top) and 9 (bottom) (blue/green/yellow/red: DNA/lipid droplets/AdipoQ/PPAR). Lower left: superblock features (lower left) that best distinguish the two selected images. (c) 2D MDS plot of PhenoRipper profiles for images of the 1820 "hits" from a genome-wide siRNA screen performed by Fuchs et al. Colors represent the six tightest phenotypic groups defined by Fuchs et al.; Class: 1-metaphase, 2-high-actin ratio, 3-lamellapodia + high-actin ratio, 4-proliferating cells, 5-small cells, 6-big cells, gray dots – other/unclassified. A.U.: arbitrary units.