

Fly-QMA: Automated analysis of mosaic imaginal discs in *Drosophila*

Sebastian Bernasek^{1,2}, Nicolás Peláez^{3,#}, Richard Carthew^{2,3,4}, Neda Bagheri^{1,2,5*}, and Luís Amaral^{1,2,5,6*}

¹ Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA

² NSF-Simons Center for Quantitative Biology, Northwestern University, Evanston, IL, USA

³ Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA

⁴ Department of Biochemistry and Molecular Genetics, Northwestern University, Evanston, IL, USA

⁵ Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA

⁶ Department of Physics and Astronomy, Northwestern University, Evanston, IL, USA

Current Address: Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. Howard Hughes Medical Institute.

* Corresponding author

E-mail: n-bagheri@northwestern.edu (NB) or amaral@northwestern.edu (LANA)

Abstract

Mosaic analysis provides a means to probe developmental processes in situ by generating loss-of-function mutants within otherwise wildtype tissues. Combining these techniques with quantitative microscopy enables researchers to rigorously compare RNA or protein expression across the resultant clones. However, visual inspection of mosaic tissues remains common in the literature because quantification demands considerable labor and computational expertise. Practitioners must segment cell membranes or cell nuclei from a tissue and annotate the clones before their data are suitable for analysis. Here, we introduce Fly-QMA, a computational framework that automates each of these tasks for confocal microscopy images of *Drosophila* imaginal discs. The framework includes an unsupervised annotation algorithm that incorporates spatial context to inform the genetic identity of each cell. We use a combination of real and synthetic validation data to survey the performance of the annotation algorithm across a broad range of conditions. By contributing our framework to the open-source software ecosystem, we aim to contribute to the current move toward automated quantitative analysis among developmental biologists.

Author summary

Biologists use mosaic tissues to compare the behavior of genetically distinct cells within an otherwise equivalent context. The ensuing analysis is often limited to qualitative insight. However, it is becoming clear that quantitative models are needed to unravel the complexities of many biological systems. In this manuscript we introduce Fly-QMA, an open-source software framework that automates the quantification of mosaic analysis for *Drosophila* imaginal discs, a common setting for studies of developmental processes. The software automatically extracts quantitative measurements from confocal images of mosaic tissues, rectifies any cross-talk between fluorescent reporters, and identifies clonally-related subpopulations of cells. Together, these functions allow users to rigorously ascribe changes in gene expression to the presence or absence of particular genes. We validate the performance of our framework using both real and synthetic data. Through its publication, we aim to contribute to the current move toward

automated quantitative analysis among biologists.

Introduction

Quantification will be essential as biologists study increasingly complex facets of organismal development [1]. Unfortunately, qualitative analysis remains common because it is often difficult to measure cellular processes in their native context. Modern fluorescent probes and microscopy techniques make such measurements possible, but the ensuing image analysis demands specialized skills that fall beyond the expertise of most experimentalists. Automated analysis strategies have addressed similar challenges in cytometry [5–7], genomics and transcriptomics [8–11], and other subdisciplines of biology [12, 13]. Image analysis has proven particularly amenable to automation, with several computer vision tools having gained traction among biologists [14–17]. These platforms are popular because they increase productivity, improve the consistency and sensitivity of measurements, and obviate the need for specialized computational proficiency [18–20]. Designing similar tools to help biologists probe and measure developmental processes in vivo will further transform studies of embryogenesis and development into quantitative endeavors.

Developmental biologists study how the expression and function of individual genes coordinate the emergence of adult phenotypes. They often ask how cells respond when a specific gene, RNA, or protein is perturbed during a particular stage of development. Cell response may be characterized by changes in morphology, or by changes in the expression of other genes (Fig 1A). Experimental efforts to answer this question were historically stifled by the difficulty of isolating perturbations to a single developmental context, as the most interesting perturbation targets often confer pleiotropic function across several stages of development and can trigger early embryonic lethality [21–23].

Mosaic analysis addressed this challenge in *Drosophila* by limiting perturbations to a subset of cells within the imaginal discs of the larva [24, 25]. The technique yields a heterogeneous tissue comprised of genetically distinct patches of cells that are clonally related. Aside from rare de novo mutations, cells within each clone are genetically identical. Clone formation may be restricted to specific developing organs by using disc-specific gene promoters to drive trans-chromosomal recombination events in the

corresponding imaginal discs [26,27]. The timing of these events determines the number and size of the resultant clones [28]. Perturbations are applied by engineering the dosage of a target gene to differ across clones (Fig 1B), resulting in clones whose cells are either homozygous mutant ($-/-$), heterozygous wildtype ($+/-$), or homozygous wildtype ($+/+$) for the particular gene. Labeling these clones with the presence or absence of fluorescent markers enables direct comparison of cells subject to control or perturbation conditions, while maintaining otherwise equivalent developmental and physiological histories between the two cell populations (Fig 2A). Additional reporters may be used to monitor differences in RNA or protein expression, morphology, or cell fate choice across clones (Fig 2B). Variants of this strategy led to seminal discoveries in both neural patterning [29–31] and morphogenesis [32,33], and remain popular today [34–36].

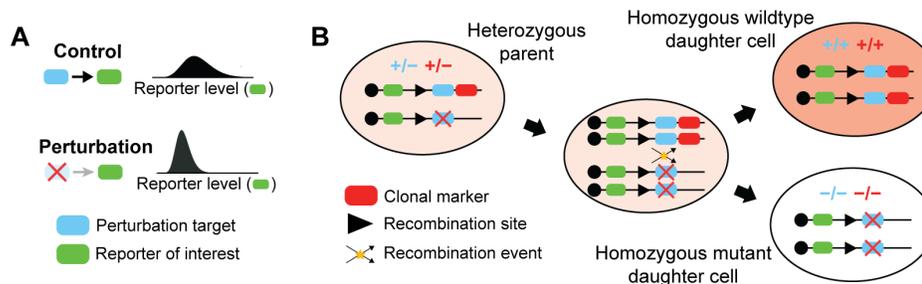


Fig 1. Perturbing gene expression via mitotic recombination. Experimental framework using mitotic clones to test whether or not regulatory interactions occur between a perturbation target and reporter of interest. Blue and green markers represent the respective genes encoding the perturbation target and the reporter. (A) A perturbation-induced decrease in reporter levels would confirm that regulation occurs. (B) Mitotic recombination generates clonal subpopulations carrying zero, one, or two copies of the gene encoding a perturbation target. Black lines depict a genetic locus. Only genes downstream of the recombination site are subject to recombination. Red markers represent a gene encoding a clonal marker used to identify the resultant clones. Red shading of large oval reflects relative clonal marker fluorescence level.

Quantitative microscopy techniques are well suited to measuring differences in cell behavior across clones. One reporter (a clonal marker) labels the clones, while others quantitatively report properties of their constituent cells, such as the expression level of a gene product of interest (Fig 2C). The former then defines the stratification under which the latter are compared. We call this strategy Quantitative Mosaic Analysis (QMA) because it replaces subjective visual comparison with a rigorous statistical alternative. Although a few recent studies have deployed this approach [37–40], qualitative visual comparison remains pervasive in the literature.

We suspect the adoption of QMA has been hindered by demand for specialized 49
computational skills or, in their stead, extensive manual labor. Researchers must first 50
draw or detect boundaries around individual nuclei in a procedure known as 51
segmentation (Fig 2D). Averaging the pixel intensities within each boundary then yields 52
a fluorescence intensity measurement for each reporter in each identified nucleus (Fig 53
2E). The measurements should then be corrected to account for any fluorescence 54
bleedthrough between reporter channels (Fig 2F). Correction often requires 55
single-reporter calibration experiments to quantify any potential crosstalk between 56
different fluorophores, followed by complex calculations to remedy the data [41, 42]. 57
Researchers must then label, or annotate, each identified nucleus as mutant, 58
heterozygous, or homozygous for the clonal marker. Annotation is typically achieved 59
through visual inspection (Fig 2G). Cells carrying zero, one, or two copies of the clonal 60
marker should exhibit low, medium, or high average levels of fluorescence, respectively. 61
However, both measurement and biological noise introduce the possibility that some 62
cells' measured fluorescence levels may not reliably reflect their genetic identity. 63
Annotation must therefore also consider the spatial context surrounding each nucleus. 64
For instance, a nucleus whose neighbors express high levels of the clonal marker is likely 65
to be homozygous for the clonal marker, even if its individual fluorescence level is 66
comparable to that of heterozygous cells (Fig 2G, white arrows). Spatial context is 67
particularly informative in developing tissues where cell migration is minimal, such as 68
the fly imaginal discs. With many biological replicates containing thousands of cells 69
each, annotation can quickly become insurmountably tedious. The corrected and 70
labeled measurements are then curated for statistical comparison by excluding those on 71
the border of each clone, and limiting their scope to particular regions of the image field 72
(Fig 2H). Combined, all of these tasks ultimately burden researchers and raise the 73
barrier for adoption of QMA. 74

Automation promises to alleviate this bottleneck, yet the literature bears 75
surprisingly few computational resources designed to support QMA. The ClonalTools 76
plugin for ImageJ deploys an image-based approach to measure macroscopic features of 77
clone morphology, but is limited to binary classification of mutant versus non-mutant 78
tissue and offers no functionality for comparing reporter expression across clones [43]. 79
Alternatively, the MosaicSuite plugin for ImageJ deploys an array of image processing, 80

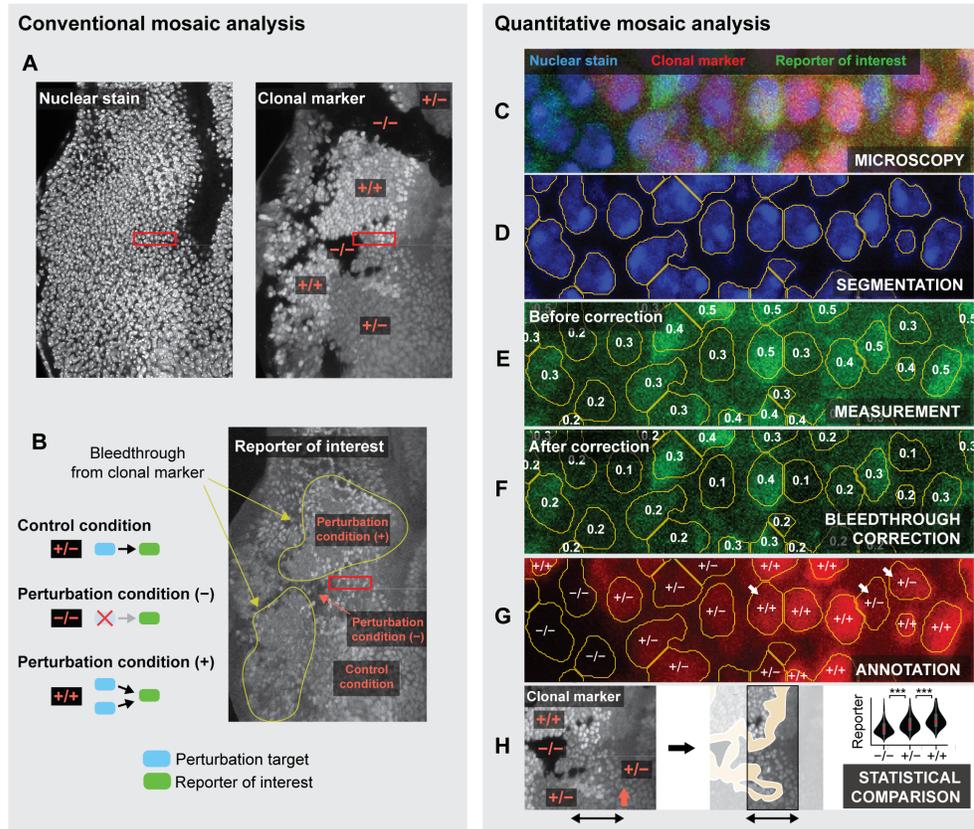


Fig 2. Conventional versus quantitative mosaic analysis. (A,B) Conventional analysis of a mosaic eye imaginal disc. (A) Clones are identified by visual comparison of clonal marker fluorescence among nuclei. (B) Regions labeled homozygous mutant (-/-) or homozygous wildtype (+/+) for the clonal marker are compared with those labeled heterozygous wildtype (+/-) to assess whether reporter expression differs across clones. Fluorescence bleed-through is arbitrarily diagnosed. (C-H) Quantitative mosaic analysis. Panels depict a magnified view of the region enclosed by red rectangles in panels A and B. (C) Raw confocal image of the nuclear stain, clonal marker, and reporter of interest. (D) Segmentation identifies distinct nuclei. (E) Reporter expression is quantified by averaging the pixel intensities within each segment. Numbers reflect measured values. (F) Measurements may be corrected to mitigate fluorescence bleedthrough. (G) Individual nuclei are labeled homozygous mutant, heterozygous, or homozygous wildtype for the clonal marker. White arrows mark nuclei with ambiguous fluorescence levels. (H) Reporter levels are compared across clones to determine whether the perturbation affects reporter expression. Yellow region marks excluded clone borders. Comparison may exclude clone borders (yellow regions) and focus on a particular region of the image field (black arrows). In the eye imaginal disc, comparison is often limited to a narrow window near the MF (orange arrow).

segmentation, and analysis capabilities to automatically detect spatial interactions 81

between objects found in separate fluorescence channels [44, 45]. While useful in many 82

other settings, neither of these tools support automated labeling of individual cells or 83

explicit comparison of clones with single-cell resolution. Most modern studies employing 84

a quantitative mosaic analysis instead report using some form of ad hoc semi-automated pipeline built upon ImageJ [37, 39, 40]. We are therefore unaware of any platforms that offer comprehensive support for an automated QMA workflow.

Here, we introduce Fly-QMA, an open-source framework for automated QMA of *Drosophila* imaginal discs. Fly-QMA supports segmentation, bleedthrough correction, and annotation of confocal microscopy data (Fig 2D-H). We demonstrate each of these functions by applying them to real confocal images of clones in the eye imaginal disc, and find that our automated approach yields results consistent with manual analysis by a human expert. We then generate and use synthetic data to survey the performance of our framework across a broad range of biologically plausible conditions.

Results

Quantification of nuclear fluorescence levels

We implemented a segmentation strategy based upon a standard watershed approach [52]. Briefly, we construct a foreground mask by Otsu thresholding the nuclear stain or nuclear label image following a series of smoothing and contrast-limited adaptive histogram equalization operations [52, 53]. We then apply a Euclidean distance transform to the foreground mask, identify the local maxima, and use them as seeds for watershed segmentation. When applied to the microscopy data, few visible spots in the nuclear stain were neglected, and the vast majority of segments outlined individual nuclei (S1 Fig C).

This approach is flexible and should perform adequately in many scenarios. However, we acknowledge that no individual strategy can address all microscopy data because segmentation is strongly context dependent. All subsequent stages of analysis were therefore designed to be compatible with any data that conform to our standardized file structure. This modular arrangement grants users the freedom to use one of the many other available segmentation platforms [54], including FlyEye Silhouette [55], before applying the remaining functionalities of our framework. Regardless of how nuclear contours are identified, averaging the pixel intensities within them yields fluorescence intensity measurements for each reporter in each identified nucleus. We next sought to

ensure that these measurements were suitable for comparison across clones. 114

Bleedthrough correction 115

Despite efforts to select non-overlapping reporter bandwidths and excite them 116
sequentially, it is not uncommon for reporters excited at one wavelength to emit some 117
fluorescence in the spectrum collected for another channel (Fig 2B, yellow lines) [41,56]. 118
The end result is a positive correlation, or crosstalk, between the measured fluorescence 119
intensities of two or more reporters. Exogenous correlations between the measured 120
fluorescence intensities of the clonal marker and the reporter of interest are problematic 121
given that the purpose of the experiment is to detect changes in reporter levels with 122
respect to the clonal marker. 123

In our microscopy data, individual clones were distinguished by their low, medium, 124
or high expression levels of an RFP-tagged clonal marker (Fig 3A). These images should 125
not have shown any detectable difference in GFP levels across clones because all cells 126
carried an equivalent dosage of the control reporter (S1 Fig A). However, the images 127
visibly suffered from bleedthrough between the RFP and GFP channels (Fig 3A,B). 128
Bleedthrough was similarly evident when we compared measured GFP levels across 129
labeled clones. Nuclei labeled mutant, heterozygous, or homozygous for the clonal 130
marker had low, medium, and high expression levels of the control reporter, respectively 131
(Fig 3C, black boxes). The data were therefore ripe for systematic correction. 132

Spectral bleedthrough correction is common practice in other forms of 133
cross-correlation and co-localization microscopy [41,56]. These methods typically entail 134
characterizing the extent of crosstalk between fluorophores globally [57,58], on a 135
pixel-by-pixel basis [42], or by experimental calibration [41], then detrending all images 136
or measurements prior to subsequent analysis. Our framework adopts the global 137
approach, using the background pixels in each image to infer the extent of fluorescence 138
bleedthrough across spectral channels. 139

Specifically, we assume the fluorescence intensity F_{ij} for channel i at pixel j is a 140
superposition of a background intensity B_{ij} and some function of the expression level 141

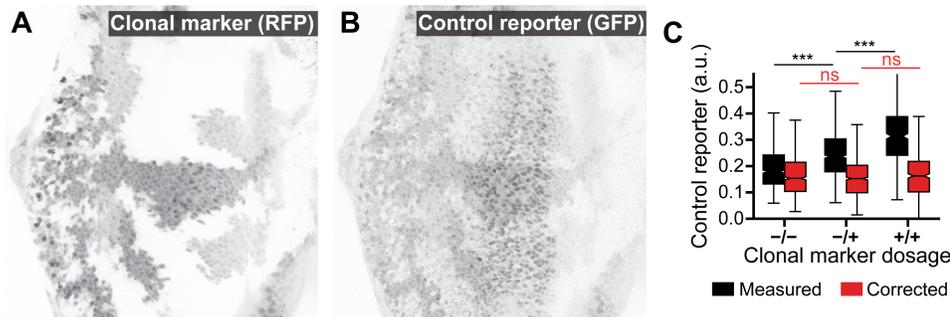


Fig 3. Automated correction of fluorescence bleedthrough in the larval eye.

(A) Low, medium, and high expression levels of the RFP-tagged clonal marker. (B) GFP-tagged control reporter expression. RFP fluorescence bleedthrough is visually apparent upon comparison with A. (C) Comparison of control reporter expression between clones. Includes data aggregated across nine images taken from six separate eye discs. Data were limited to cells within the region of elevated GFP expression that were of approximately comparable developmental age (see S2 Fig E-G). Measurements are stratified by their assigned labels. Before correction, expression differs between clones (black boxes, $p < 10^{-5}$). No difference is detected after correction (red boxes, $p > 0.05$).

E_{ij} that we seek to compare across cells [59]:

142

$$F_{ij} = B_{ij} + f(E_{ij}) \quad (1)$$

We further assume that the background intensity of a channel includes linear

143

contributions from the fluorescence intensity of each of the other channels:

144

$$B_{ij} = \sum_{k \neq i} \alpha_k F_{kj} + \beta \quad (2)$$

where k is indexed over K anticipated sources of bleedthrough. Given estimates for each

145

$\{\alpha_1, \alpha_2, \dots, \alpha_K\}$ and β we can then estimate the background intensity of each

146

measurement:

147

$$\langle B_{ij} \rangle = \sum_{k \neq i} \alpha_k \langle F_{kj} \rangle + \beta \quad (3)$$

where the braces denote the average across all pixels within a single nucleus. The

148

corrected signal value is obtained by subtracting the background intensity from the

149

measured fluorescence level:

150

$$\langle f(E_{ij}) \rangle = \langle F_{ij} \rangle - \langle B_{ij} \rangle \quad (4)$$

Repeating this procedure for each nucleus facilitates comparison of relative
expression levels across nuclei in the absence of bleedthrough effects. Bleedthrough
correction performance is therefore strongly dependent upon accurate estimation of the
bleedthrough contribution strengths, $\{\alpha_1, \alpha_2, \dots, \alpha_K\}$.

We estimate these parameters by characterizing their impact on background pixels
(see Methods). When applied to the microscopy data, bleedthrough correction
successfully eliminated any detectable difference in GFP expression across clones (Fig
2C, red boxes, $p > 0.05$ two-sided Mann-Whitney U test).

Automated annotation of clones

Our annotation strategy seeks to label each identified cell as homozygous mutant,
heterozygous wildtype, or homozygous wildtype for the clonal marker. Variation within
each clone precludes accurate classification of a cell's genotype solely on the basis of its
individual expression level. However, in tissues where cell migration is minimal, clonal
lineages are unlikely to exist in isolation because recombination events are typically
timed to generate large clones. Our strategy therefore integrates both clonal marker
expression and spatial context to identify clusters of cells with locally homogeneous
expression behavior, then maps each cluster to one of the possible labels. This
unsupervised approach lends itself to automated annotation because the clusters are
inferred directly from the data without any guidance from the user.

We first train a statistical model to estimate the probability that a given
measurement came from a cell carrying zero, one, or two copies of the clonal marker (S3
Fig A). This entails fitting a weighted mixture of three or more bivariate lognormal
distributions (components) to a two dimensional set of observations (S3 Fig B,C). The
first dimension corresponds to the clonal marker fluorescence level measured within each
cell. The second dimension describes the local average expression level within the region
surrounding each cell. We evaluate the latter by estimating a neighborhood radius from
the decay of the radial correlation of the expression levels, then averaging the expression
levels of all cells within that radius (S3 Fig D). The second dimension therefore
measures the spatial context in which a cell resides. We balance model fidelity against
overfitting by using the Bayesian information criterion to determine the optimal number

of model components (S3 Fig E). We then cluster the components into three groups on 181
the basis of their mean values (S3 Fig F), effectively mapping each component to one of 182
the three possible gene dosages. The model may be trained using observations derived 183
from a single image, or with a collection of observations derived from multiple images. 184
Once trained, the model is able to predict the conditional probability that an individual 185
observation belongs to one of the model's components, given its measured expression 186
level. 187

We then use the learned conditional probabilities to detect entire clones, thus 188
assigning a label to each cell. Rather than using the trained model to classify each 189
observation, we compile a new set of observations by limiting each estimate of spatial 190
context to spatially collocated communities with similar expression behavior (S4 Fig A). 191
We identify these communities by applying a community detection algorithm to an 192
undirected graph connecting adjacent cells (S4 Fig B). Edges in this graph are weighted 193
by the similarity of clonal marker expression between neighbors, resulting in 194
communities with similar expression levels (S4 Fig E, Steps I and II). The graph-based 195
approach increases spatial resolution by limiting the information shared by dissimilar 196
neighbors. Applying the mixture model yields an initial estimate of the probability that 197
an observation belongs to one of the model's components (S4 Fig E, Step III). We 198
further refine these estimates by allowing the probabilities estimated for each cell to 199
diffuse throughout the graph (S4 Fig E, Step IV). The rate of diffusion between 200
neighbors is determined by the weight of the edge that connects them, with more 201
similar neighbors exerting stronger influence on each other. We then use the diffused 202
probabilities to identify the most probable source component and label each observation 203
(S4 Fig E, Step V). These probabilities also provide a measure of confidence in the 204
assigned labels. We replace any low-confidence labels with alternate labels assigned 205
using a marginal classifier that neglects spatial context (S4 Fig F,G), resulting in a fully 206
labeled image (S4 Fig H). 207

The algorithm leverages the collective wisdom of neighboring measurements to 208
override spatially isolated fluctuations in clonal marker expression, and thereby enforces 209
consistent annotation within contiguous regions of the image field. The size of these 210
regions depends upon the granularity of estimates for the spatial context surrounding 211
each cell. We used an unsupervised approach to choose an appropriate spatial resolution 212

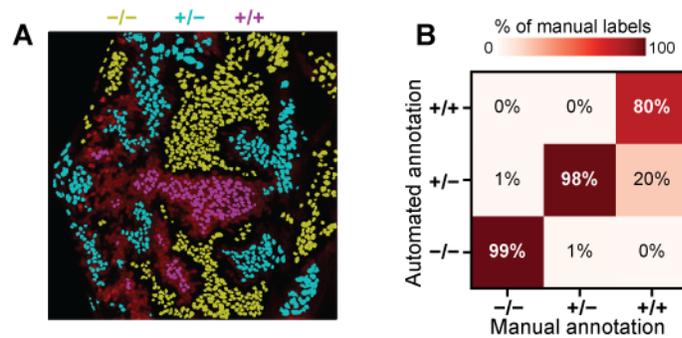


Fig 4. Automated unsupervised annotation of clones in the larval eye. (A) Labels assigned by automated annotation. Yellow, cyan, and magenta denote the label assigned to each contour. Labels are overlaid on the RFP channel of the image shown in S1 Fig B. Cells on the periphery of each clone are excluded. (B) Comparison of automated annotation with manually-assigned labels. Confusion matrix includes data aggregated across nine images taken from six separate eye discs. Cells on the periphery of each clone are included. Columns sum to one.

in a principled manner. In short, the resolution is matched to the approximate length 213
scale over which expression levels remain correlated among cells. Both the training and 214
application stages of our annotation algorithm use this automated approach (S3 Fig D 215
and S4 Fig D), thus averting any need for user input. 216

Manual assessment of annotation performance 217

We sought to validate the performance of the annotation algorithm by assessing its 218
ability to accurately reproduce human-assigned labels. We manually labeled nuclei in 219
each eye imaginal disc as homozygous mutant, heterozygous wildtype, or homozygous 220
wildtype for the clonal marker, then automatically labeled the same cells (Fig 4A). The 221
two sets of labels showed strong overall agreement (Fig 4B and S5 Fig A). Excluding 222
cells on the border of each clone revealed greater than 97% agreement in seven of the 223
nine annotated images (see Table 1). Upon secondary inspection of the sole instance of 224
substantial disagreement (S5 Fig B), we are unable to confidently discern which set of 225
labels are more accurate. 226

While it is common practice to use human-labeled data as the gold standard, 227
manually assigned labels do not represent a reliable and reproducible ground truth. 228
Furthermore, we contend that validation with manually-labeled data entrains implicit 229
human biases in the selection of performant algorithms. These biases are particularly 230

Table 1. Automated vs. manual annotation

Disc	Layer	Agreement*
1	1	93.1% (97.3%)
1	2	95.3% (97.3%)
2	1	91.3% (99.1%)
2	2	95.2% (96.4%)
3	1	67.2% (75.6%)
4	1	82.5% (89.2%)
5	1	96.2% (100%)
6	1	99.1% (99.3%)
6	2	95.2% (97.5%)

*Values in parentheses denote agreement when clone borders are excluded.

pronounced in biological image data where intrinsic variation, measurement noise, and
transient processes can make cell-type annotation a highly subjective, and thus
irreproducible, task.

Synthetic benchmarking of annotation performance

Synthetic benchmarking provides a powerful alternative to validation against manually
labeled data. The idea is simple; measure how accurately an algorithm is able to label
synthetic data for which the labels are known. The synthetic data generation procedure
may be modeled after the process underlying formation of the real data, providing a
means to assess the performance of an algorithm across the range of conditions that it is
likely to encounter. The strategy therefore provides a means to survey the breadth of
biologically plausible conditions under which the algorithm provides adequate
performance. Synthetic benchmarking also facilitates unbiased comparison of competing
algorithms, resulting in a reliable standard that may be called upon at any time.

We used synthetic microscopy data to benchmark the performance of our annotation
strategy. Each synthetic dataset depicts a simulated culture of cells distributed roughly
uniformly in space (S6 Fig A). Cells in this culture contain zero, one, or two copies of a
gene encoding an RFP-tagged clonal marker (S6 Fig B). Our simulation procedure
ensures that cells tend to remain proximal to their clonal siblings (S6 Fig C), thus
forming synthetic clones with tunable size and spatial heterogeneity (S6 Fig D,E). We
generated synthetic measurements by randomly sampling fluorescence levels in a
dosage-depend manner (S7 Fig A-C). We varied the similarity of fluorescence levels

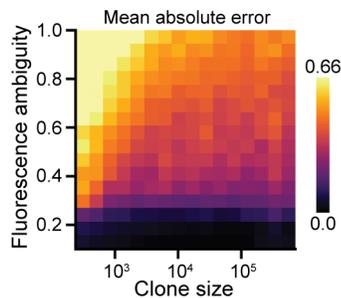


Fig 5. Synthetic benchmarking of automated annotation performance. Each pixel reflects the mean MAE across 50 replicates. Clone size reflects the mean number of cells per clone. Performance improves with increasing clone size and worsens with increasing fluuquity.

across clones using an ambiguity parameter, σ_α , that modulates the spread of the 252
distributions used to generate fluorescence levels (S7 Fig D-F). 253

Using this schema as a template, we generated a large synthetic dataset, annotated 254
each set of measurements, and compared the assigned labels with their true values. We 255
used the mean absolute error as a comparison metric because it provides a stable 256
measure of accuracy for multiclass classification problems in which the labels are 257
intrinsically ordered [60]. In other words, it penalizes egregious misclassifications more 258
severely than mild ones. 259

Annotation performance is very strong for all cases in which $\sigma_\alpha \leq 0.3$ (Fig 5). 260
Unsurprisingly, performance suffers as the difficulty of the classification problem is 261
increased. The same trends are evident when performance is graded strictly on accuracy 262
(S8 Fig). As cells on the periphery of each clone were not excluded from these analyses, 263
the observed metrics provide a lower bound on the performance that may be anticipated 264
in practice. 265

Performance improved with increasing clone size. We suspected this was caused by 266
larger clones offering additional spatial context to inform the identify of each cell. We 267
verified our assertion by re-evaluating performance relative to a variant of our 268
annotation algorithm that neglects spatial context (S4 Fig G). As expected, the 269
variant's performance exhibited no dependence on clone size (S9 Fig A). Comparing the 270
two strategies confirmed that spatial context confers the most benefit when clones are 271
large (S9 Fig B). Inclusion of spatial context also becomes increasingly advantageous as 272
the fluorescence ambiguity is increased, even for smaller clones. Thus, spatial context 273

adds progressively more value as the classification task becomes more difficult. 274

This observation may be rationalized from a statistical perspective. Each cell is 275
classified by maximizing the probability that the assigned label is correct. We compute 276
these probabilities using the estimated expression level of each cell. Neglecting spatial 277
context, this estimate is limited to a single sample and is therefore highly sensitive to 278
both measurement and biological noise. Incorporating spatial context expands the 279
sample size and thereby reduces the standard error of the estimated fluorescence level. 280
The strategy is thus generally well suited to scenarios in which fluorescence intensities 281
correlate across large clones, and closely parallels computer vision methods that exploit 282
spatial contiguity to segment image features with ill-defined borders [61]. Because 283
increased measurement precision comes at the expense of spatial resolution, we expect 284
strong performance when measurements are aggregated across relatively large clones, 285
but failure to detect small, heterogeneous clones. These expectations are consistent with 286
the observed results. They are also conveniently aligned with the anticipated properties 287
of real data, as experiments typically attempt to mitigate edge effects by driving early 288
recombination events to generate large clones. 289

Conclusion 290

We used synthetic data to survey the performance of our annotation strategy across a 291
much broader range of conditions than would have otherwise been possible with 292
manually labeled data. This included conditions well beyond those of practical use. In 293
particular, experiments designed to compare gene expression levels across clones would 294
likely seek to avoid generating small clones with ambiguous clonal marker expression. 295
Beyond complicating the annotation task, small clones are also exposed to 296
diffusion-mediated signals from adjacent clones that can mask the effect of mutations. 297
Cells located near the clone boundaries are often excluded for the same reason, as 298
quantification is typically most reliable in cells surrounded by similar neighbors. 299
Synthetic data provided a means to survey these edge cases and establish a lower bound 300
on annotation performance. The strong performance observed across the remaining 301
conditions bolsters our confidence that our annotation strategy is well suited to the 302
images it is likely to encounter. 303

In each of our examples, clones were distinguished by ternary segregation of clonal marker fluorescence levels. Modern mosaic analysis techniques continue to deploy ternary labeling [62,63], but also frequently opt for binary labeling of mutant versus non-mutant clones [64–66] and dichromic labeling of twin-spots [67,68]. Our annotation scheme readily adapts to each of these scenarios provided that the number of anticipated labels is adjusted accordingly. In the case of dichromic labeling, binary classification would be performed separately for each color channel before merging the assigned labels. Extending the same logic to combinatorial pairs of colors suggests that our framework may also be compatible with multicolor labeling schemes used to simultaneously trace many clonal lineages over time [69–71]. Our framework is thus well suited to many different mosaic analysis platforms deployed in imaginal discs.

In principle, the framework described here should also be applicable to a wide variety of other tissues [72,73] and model organisms [74–76] in which mosaics are studied. In practice, application to alternate contexts would require modifying some stages of the analysis. Most notably, image segmentation is strongly context dependent and any attempts to develop a universally successful strategy are likely to prove futile [77]. For this reason, we implemented a modular design in which each stage of analysis may be applied separately. For example, a user could perform their own segmentation before using our bleedthrough correction and clone annotation tools. By offering modular functionalities we hope to extend the utility of our software to the wider community of developmental biologists. Furthermore, the open-source nature of our framework supports continued development of more advanced features as various demands arise. Our synthetic benchmarking platform could then be used to objectively confirm the benefit conferred by any future developments.

Materials and methods

Genetics and microscopy of *Drosophila* eye imaginal discs

We borrowed an experimental dataset from a separate study of neuronal fate commitment during eye disc development [38]. The data consist of six eye imaginal discs dissected and fixed during the third larval instar of *Drosophila* development. Within

each disc, *ey>FLP* and *FRT40A* were used to generate clones. The chromosome arm 333
(2L) targeted for recombination was marked with a *Ubi-mRFPnls* transgene (S1 Fig A), 334
enabling automated detection of clones marked by distinct levels of mRFP fluorescence 335
(S1 Fig B). The discs also carried a *pnt-GFP* reporter transgene located on a different 336
chromosome that was not subject to mitotic recombination. Discs were dissected, fixed, 337
and co-stained with a DAPI nuclear dye prior to confocal imaging. Please refer to the 338
original study for additional details regarding genetics and experimental conditions. 339

The PntGFP reporter is predominantly expressed in two narrow stripes of progenitor 340
cells during eye disc development [38]. The first stripe occurs immediately posterior to a 341
wave of developmental signaling that traverses the eye disc. Progenitor cells located in 342
this region are suitable for comparison because they are of approximately equivalent 343
developmental age. We applied the Fly-QMA framework to a total of nine images of 344
these cells. 345

Characterization of fluorescence bleedthrough 346

For each image, we morphologically dilate the foreground until no features remain 347
visible (S2 Fig A). We then extract the background pixels and resample them such that 348
the distribution of pixel intensities is approximately uniform (S2 Fig B). Resampling 349
helps mitigate the skewed distribution of pixel intensities found in the background. We 350
then estimate values for each $\{\alpha_1, \alpha_2, \dots, \alpha_K\}$ and β by fitting a generalized linear 351
model to the fluorescence intensities of the resampled pixels (S2 Fig C). Each model is a 352
variant of Eq 3 in which angled braces instead denote averages across all background 353
pixels. We formulate these models with identity link functions under the assumption 354
that residuals are gamma distributed. Their coefficients provide an estimate of the 355
bleedthrough contribution strengths that may then be used to estimate the background 356
fluorescence intensity of each nucleus in the corresponding image (S2 Fig D). The 357
measurements may then be corrected through application of Eq 4. 358

Clone annotation algorithm

359

We assume the measured fluorescence level x_i for cell i is sampled from an underlying distribution $p_m(x)$ for cells carrying m copies of the gene encoding the clonal marker:

360

361

$$x_i \sim p_m(x) \quad (5)$$

We further assume that $p_m(x)$ is comprised of a mixture of one or more lognormal distributions:

362

363

$$p_m(\ln x) = \sum_{n=1}^N \lambda_n \mathcal{N}(\ln x | \theta_n) \quad (6)$$

$$\sum_{n=1}^N \lambda_n = 1 \quad (7)$$

where $0 \leq \lambda \leq 1$ are the mixing proportions, $\theta_n = (\mu_n, \sigma_n^2)$ are the mean and variance of the n th distribution. This assumption is supported by both empirical observations and theoretical insights [46, 47]. By superposition, the global distribution of measured fluorescence levels $p(\ln x)$ for all values of m are also sampled from a mixture of K components:

364

365

366

367

368

$$p(\ln x) = \sum_{m=0}^2 \alpha_m p_m(\ln x) = \sum_{m=0}^2 \alpha_m \sum_{n=1}^N \lambda_n \mathcal{N}(\ln x | \theta_n) = \sum_{k=1}^K \lambda_k \mathcal{N}(\ln x | \theta_k) \quad (8)$$

$$\sum_{k=1}^K \lambda_k = 1 \quad (9)$$

where α_m denotes the overall fraction of cells with m copies of the gene encoding the clonal marker. For brevity, we substitute $X = \ln x$ yielding:

369

370

$$p(X) = \sum_{k=1}^K \lambda_k \mathcal{N}(X | \theta_k) \quad (10)$$

Given a collection of sampled fluorescence levels, $\{X_i\}_{i=1 \dots N}$, we use expectation maximization to find values of θ_k and λ_k for each of the model's K components that maximize the log-likelihood of the observed sample. We repeat this procedure for a range of sequential values of K , resulting in multiple models of increasing size. We then balance model resolution against overfitting by selecting the model that yields the

371

372

373

374

375

smallest value of the Bayesian Information Criterion (BIC):

376

$$BIC(K) = \ln(N)q_K - 2\ln(\hat{L}_K) \quad (11)$$

$$q_K = K - 1 + 2^K \quad (12)$$

where N is the sample size, $\ln(\hat{L})_K$ is the maximum value of the log-likelihood, the subscript K denotes the number of mixture components in the model, and q_K is the total number of parameters (i.e. $K - 1$ values of λ_k and 2^K values of μ_k and σ_k^2).

377

378

379

Applying Bayes' rule to the selected model infers the posterior probabilities that each sample X_i belongs to the k th component:

380

381

$$p(k|X_i) = \frac{p(X_i|k)p(k)}{p(X_i)} = \frac{p(X_i|k)\lambda_k}{p(X_i)} \quad (13)$$

where $p(X_i | k)$ is evaluated using the model's likelihood function and $p(X_i)$ is evaluated by marginalizing across each of the model's K components. The end result is a mixture model that allows us to predict the probability that a given measurement of clonal marker expression belongs to a particular one of its component distributions.

382

383

384

385

We then define a many-to-one mapping, f , from each of the K components of the mixture to each of the three possible values of m :

386

387

$$f : \{0, 1, \dots, K\} \rightarrow \{0, 1, 2\} \quad (14)$$

We determine the mapping by k-means clustering the K component distributions into three groups on the basis of their mean values, e^{μ_k} . We may then assign a genotype label m to each measurement X_i by predicting the component k from which it was sampled.

388

389

390

391

The accuracy of these labels depends upon how closely the fitted mixture model reflects the true partitioning of gene copies among clones. While finite mixtures are always identifiable given a sufficiently large sample [48], the algorithm used to fit the mixture tends toward local maxima of the likelihood function when the true components are similar (Wu, 1983). An approach based on a univariate mixture is thus inherently prone to failure when expression levels extensively overlap across clones, as

392

393

394

395

396

397

variation within each clone precludes accurate classification of a cell's genotype solely on 398
the basis of its individual expression level. However, clonal lineages are unlikely to exist 399
in isolation because recombination events are usually timed to generate large clones. 400
Our strategy therefore integrates both clonal marker expression and spatial context to 401
identify clusters of cells with locally homogeneous expression behavior. 402

We incorporate spatial context by introducing a second jointly-distributed variable 403
 Y_i : 404

$$Y_i = \frac{1}{M_i} \sum_{j=0}^{M_i} X_j \quad (15)$$

where the subscript j indexes all M_i neighbors of cell i . The new variable reflects the 405
average expression level among the neighbors surrounding each cell. We define 406
neighbors as pairs of cells located within a critical distance of each other. This distance, 407
or sampling radius, is derived from the approximate length scale over which cells retain 408
approximately similar clonal marker expression levels. Specifically, we determine the 409
exponential decay constant of the spatial correlation function, $\psi(\delta)$: 410

$$\psi(\delta) = \frac{\langle (X_i - \mu_X)(X_j - \mu_X) \rangle_{i,j \in \delta}}{\sigma_X^2} \quad (16)$$

where μ_X and σ_X^2 are the global mean and standard deviation, and angled brackets 411
denote the mean across all pairs of cells separated by distance δ . We efficiently 412
implement this procedure by fitting an exponential decay function to the down-sampled 413
moving average of $\psi(\delta)$ as a function of increasing separation distance. 414

Following the introduction of spatial context, the mixture model becomes: 415

$$p(X, Y) = \sum_{k=1}^K \lambda_k \mathcal{N}(X, Y | \theta_k) \quad (17)$$

where $\theta_k = (\vec{\mu}_k, \vec{\sigma}_k^2)$ contains the mean and variance of each component given by vectors 416
of length two. This formulation constrains each component's covariance matrix to be 417
diagonal. The posterior is now: 418

$$p(k|X_i, Y_i) = \frac{p(X_i, Y_i|k)\lambda_k}{p(X_i, Y_i)} \quad (18)$$

We can recover the univariate model by marginalizing the posterior over all values of Y : 419

$$p(k|X_i) = \sum_j p(k|X_i, Y_j) \quad (19)$$

When neglecting spatial context, we use this expression to classify each sample by 420

applying the mapping f to the value of k that maximizes $p(k | X_i)$: 421

$$f(\operatorname{argmax}_k p(k|X_i)) \quad (20)$$

In all other cases, we deploy a graph-based approach to refine the estimate of 422

$p(k | X_i, Y_i)$. This first entails constructing an undirected graph connecting adjacent 423

cells within each image. We obtain the graph's edges through Delaunay triangulation of 424

the measured cell positions, then exclude distant neighbors by thresholding the edge 425

lengths. Each edge is assigned a weight w_{ij} reflecting the similarity of clonal marker 426

expression between adjacent cells i and j : 427

$$w_{ij} = \exp\left(\frac{-E_{ij}}{\langle E \rangle}\right) \quad (21)$$

$$E_{ij} = |X_i - X_j| \quad (22)$$

where E_{ij} is the absolute log fold-change in measured expression level and angled 428

brackets denote the mean across all edges. We chose an exponential formulation because 429

it yields an approximately uniform distribution of edge weights. We then detect 430

communities within the graph using the Infomap algorithm [49]. The algorithm provides 431

a hierarchical partitioning of nodes into non-overlapping clusters. We aggregate all 432

clusters below a critical level that is again chosen by estimating the spatial correlation 433

decay constant. We then enumerate $p(k | X_i, Y_i^c)$ where Y_i^c is the spatial context 434

obtained by averaging expression levels among all neighbors in the same community as 435

cell i . 436

We further incorporate spatial context by allowing the posterior probabilities 437

$p(k | X_i, Y_i^c)$ to diffuse among adjacent cells. We define the modified posterior 438

probability $\hat{p}(k | X_i, Y_i^c)$ through a recursive relation analogous to the Katz 439

centrality [50], initialized by $p(k | X_i, Y_i^c)$:

440

$$\hat{p}(k | X_i, Y_i^c) = \alpha \sum_j w_{ij} \hat{p}(k | X_i, Y_i^c) + \beta \quad (23)$$

$$\beta = (1 - \alpha)p(k|X_i, Y_i^c) \quad (24)$$

where α is the attenuation factor and w_{ij} are the edge weights. Expressed in matrix

441

form, the solution for $\hat{p}(k | X, Y^c)$ is given by:

442

$$\hat{p}(k | X, Y^c) = (I - \alpha W)^{-1}(1 - \alpha)p(k | X, Y^c) \quad (25)$$

where I denotes the identity matrix and W is the matrix of edge weights w_{ij} . We then

443

assign a label to each measurement X_i by applying f to the value of k that maximizes

444

$\hat{p}(k | X_i, Y_i^c)$:

445

$$f(\operatorname{argmax}_k \hat{p}(k|X_i, Y_i^c)) \quad (26)$$

Finally, we assess the total posterior probability of each assigned label, $\hat{P}(m_i)$:

446

$$\hat{P}(m_i) = \sum_{\{k|f(k)=m_i\}} \hat{p}(k|X_i, Y_i^c) \quad (27)$$

This measure reflects the overall confidence that m_i is the appropriate label. Labels

447

whose confidence falls below 80% are replaced by their counterparts estimated using the

448

marginal classifier. This substitution helps preserve classification accuracy in situations

449

where spatial context is not informative, and is particularly useful when the annotated

450

clones are relatively small.

451

Statistical comparison of fluorescence levels

452

To mitigate edge effects, cells residing on the periphery of each clone were excluded

453

from all comparisons (S2 Fig E). Border cells were identified by using a Delaunay

454

triangulation to find all cells connected to a neighbor within a different clone. Our

455

framework includes a simple graphical user interface that permits manual curation of

456

which regions of the image field are included in subsequent analyses. We used this tool

457

to limit our analysis to the region of elevated GFP expression near the morphogenetic furrow (S2 Fig F). Comparisons were further restricted to cells undergoing similar stages of development (S2 Fig G). These restrictions served to buffer against differences in developmental context and ensured that all compared cells were of similar developmental age. The remaining fluorescence measurements were then aggregated across all eye discs and compared between pairs of clones by two-sided Mann-Whitney U test.

Simulated cell growth and recombination

We simulated the two dimensional growth of a cell culture seeded with a single cell. Growth proceeds through sequential division of cells (S6 Fig A). Not all cells divide at each time-step because cell division is a stochastic process. Instead, each cell divides stochastically with a rate controlled by a global growth rate parameter.

Cells in this culture carry a gene encoding a clonal marker (S6 Fig B). During growth, the gene is subject to mitotic recombination (S6 Fig C). Each time a cell divides, its genes are duplicated and equally partitioned between the two daughter cells. However, in some instances a heterozygous parent may instead partition its two duplicate genes unequally, with one daughter receiving both and the other receiving none. These mitotic recombination events occur stochastically with a frequency defined by a global recombination rate parameter.

After each round of cell division, all cells are repositioned in order to preserve approximately uniform spatial density (S6 Fig C). Repositioning is achieved by equilibrating a network of springs connecting each cell with its neighbors. This undirected network is constructed through Delaunay triangulation of all cells spatial positions. Edges on the periphery of the culture are systematically excluded by establishing a maximum polar angle between neighbors. This filtration removes spurious edges between distant pairs of cells. Edges connecting pairs of cells with the same clonal marker dosage are assigned a 10% higher spring constant than edges that connect dissimilar cells. This modest bias ensures that cells tend to remain proximal to their clonal lineages. Cell positions are then updated using a force-directed graph drawing algorithm [51]. Alternating cell division and repositioning steps are then repeated until a predefined population size is reached.

The timing and duration of recombination events affects the number and size of the resultant clones. In real experiments, recombination events are restricted to a particular stage of the developmental program through localized exogenous expression of the recombination machinery. We incorporated this feature into our cell growth simulations via two adjustable parameters. The first determines the minimum population size at which recombination may begin, while the second determines the number of generations over which recombination may continue to occur. These two parameters provide a means to tune the average number and size of clonal subpopulations in the synthetic data (S6 Fig D). Early recombination events generally entail larger clones, while shorter recombination periods limit the extent of clone formation (S6 Fig E).

Generation of synthetic microscopy data

Each simulation yields a list of spatial coordinates and gene dosages for each nucleus (S6 Fig B). Synthetic measurements for each nucleus were generated by randomly sampling fluorescence levels $\{x_1, x_2, \dots, x_{i=N}\}$ from a lognormal distribution conditioned upon the corresponding gene dosage (S7 Fig A-C):

$$\ln x \sim \mathcal{N}_n(\theta_n) \quad (28)$$

where the subscript n denotes the gene copy number and $\theta_n = (\mu_n, \sigma_\alpha^2)$ are the mean and variance of the corresponding distribution. We define μ_n such that the mean fluorescence level doubles for each additional copy of the gene:

$$\mu_n = \ln(2^{n-1}) \quad (29)$$

We refer to σ_α as the *fluorescence ambiguity* because it modulates the similarity of fluorescence levels across gene dosages. Increasing σ_α increases the overlap among \mathcal{N}_0 , \mathcal{N}_1 , and \mathcal{N}_2 (S7 Fig D,E), and consequently increases the difficulty of the annotation task (S7 Fig F).

Synthetic benchmarking of annotation performance

510

We generated a large synthetic dataset spanning a broad range of sixteen different clone sizes and fluorescence ambiguities (S6 Fig D and S7 Fig F, only half are shown). We performed 50 replicate simulations for each condition. All simulations were terminated when the total population exceeded 2048 cells. We assigned each cell a 20% probability of division upon each iteration, and each cell division event was accompanied by a 20% chance of mitotic recombination. Parent cells containing zero or two copies of the recombined genes were ineligible for recombination, effectively sealing the genetic fates of their respective lineages.

511

512

513

514

515

516

517

518

To annotate each set of measurements, the mixture model given by Eq 17 was independently trained and applied to each replicate. Training a single model on all replicates yields modestly stronger performance on average (not shown), but also yields more variable results across the parameter space because all labels are dependent upon the outcome of a single expectation maximization routine.

519

520

521

522

523

Data and software availability

524

We have distributed the automated mosaic analysis framework as an open-source python package available at <https://github.com/sebastianbernasek/flyqma>. We also intend to incorporate its core features into future versions of *FlyEye Silhouette*, our open-source platform for quantitative analysis of the larval eye. The code used to generate synthetic microscopy data is also freely available at <https://github.com/sebastianbernasek/growth>. All segmented and annotated eye discs are accessible via our data repository (<https://doi.org/10.21985/N2F207>).

525

526

527

528

529

530

531

Supporting information

532

S1 Fig Example clones in the larval fly eye. (A) Genetic schema for a bleedthrough control experiment. Red and green ovals represent genes encoding a RFP-tagged clonal marker and a GFP-tagged control reporter, respectively. Black lines depict a genomic locus. Recombination does not affect gene dosage of the control reporter, so GFP variation across clones is attributed to fluorescence bleedthrough. (B) Confocal image of an eye imaginal disc. Red, green, and blue reflect clonal marker, control reporter, and nuclear stain fluorescence, respectively. (C) Segmentation of the DAPI nuclear stain. White lines show individual segments.

533

534

535

536

537

538

539

540

S2 Fig Using background pixels to characterize bleedthrough contributions in the foreground. (A) Extraction of background pixels (striped region). Foreground includes the merged RFP and GFP images, surrounded by a white line. White arrow marks the morphogenetic furrow (MF). (B) Background pixel values are resampled such that RFP intensities are uniformly distributed. (C) A generalized linear model characterizes the contribution of RFP bleedthrough to GFP fluorescence. Boxes reflect windowed distributions of resampled background pixel intensities. Red line shows the model fit. (D) Measured GFP levels before bleedthrough correction. Markers represent individual nuclei. Red line shows the inferred contributions of RFP fluorescence bleedthrough. Dashed portion is extrapolated. (E-G) Data curation prior to statistical comparison of GFP levels. (E) Cells on the periphery of each clone are excluded. (F) The selection is limited to the region of elevated GFP expression near the MF. (G) It is further limited to cells of the same developmental age, defined by their relative positions along the x-axis.

S3 Fig Training a clone annotation model. (A) One or more images are segmented, yielding a set of fluorescence measurements X . These are used to sample the spatial context Y of the neighborhood surrounding each cell. Both sets of values are used to train a mixture model. Subsequent panels demonstrate these procedures using the example shown in S3 Fig C. (B) Expression levels are jointly distributed with the local average among neighboring cells. Center panel shows the joint distribution. Top and right bar plots show marginal distributions. (C) Mixture model identifies seven distinct components k_i . Center panel shows position and spread of each component. Top and right panels show marginal components scaled by their respective weights. Red shading denotes the label m_i assigned to each component. The model predicts the posterior probabilities that a given sample (X, Y) belongs to each component. (D) Neighborhood size is estimated by computing the decay constant of the spatial correlation function, $\psi(\delta)$. Black line shows the moving average of $\psi(\delta)$, red line shows an exponential fit. Inset shows the resultant sampling region. (E) The optimal number of mixture components is determined by minimizing BIC score. (F) Mixture components are labeled by k-means clustering their mean values. Markers reflect the component means, colors denote the assigned label.

S4 Fig Label assignment using a trained clone annotation model. (A) Measurements are used to sample spatial contexts before the trained model is applied (blue and green path). In parallel, measurements are labeled using a marginal projection of the trained model (magenta path). The labels are then merged (red path). (B-D) Spatial context sampling. (B) Weighted undirected graph connecting adjacent cells. Line width reflects expression similarity between neighbors. (C) Community resolution is defined by aggregating clusters that fall below a hierarchical cut level δ . Panels show increasing levels of aggregation. Colors denote distinct communities. (D) Cut level is chosen by finding the maximum level (red dot) that remains lower than the decay constant of the spatial correlation function, $\psi(\delta)$ (black line). Panel E depicts aggregation below the third level for ease of visualization. (E) Application of the mixture model. (I) The graph contains distinct communities of locally similar expression. (II) Mean expression level within each community serves as the local average for each cell. (III) Mixture model estimates the probability that each cell belongs to each of its component. Bar plots within each cell illustrate the cumulative probability of each label. (IV) Posterior probabilities are diffused across the graph. (V) Each cell is assigned the most probable label. (F,G) Application of a marginal mixture model. (F) Marginal mixture components, shaded by their mapped labels. Dashed line is the overall marginal density. (G) Marginal classifier labels cells strictly on the basis of

their individual fluorescence level. Red shading denotes the most probable label for each level. (H) Annotated measurements. Red shading denotes the assigned label. Labels with low confidence $\hat{P}(m_i) < 0.8$ are replaced by their marginal counterparts.

S5 Fig Comparison of automated annotation with manually assigned labels. (A) Distribution of labels among each possible value. (B) Visual comparison of the sole instance in which automated and manual annotation differ. Image shows clonal marker fluorescence, colors denote the assigned label.

S6 Fig Simulated growth of a synthetic cell culture. (A) Partial simulation time course. Each marker depicts a cell. Greyscale intensity reflects clonal marker gene dosage. Simulation time reflects the approximate number of cell divisions since the initial seed. (B) Simulations yield gene dosages and spatial coordinates for each cell. (C) Single iteration of an example simulation. Circles represent individual cells, red shading denotes clonal marker dosage. Cycles of cell division, recombination, and repositioning are repeated until the simulation reaches a specified end time ($t > 11$ in panel A). (D) Cultures simulated with varying recombination start times. All cultures were subject to four generations of recombination ($\delta t = 4$). Recombination start time increases from left to right. Later recombination events generally yield smaller clones. (E) Mean clone size (cells per clone) as a function of the recombination start time. Colors denote recombination period duration. Error bars reflect standard error of the mean across 50 replicates. Clone size generally decreases as recombination is limited to later times.

S7 Fig Tunable generation of synthetic microscopy data. (A) Fluorescence levels are sampled from lognormal distributions conditioned upon gene dosage. (B) Synthetic data include a measured fluorescence level for each reporter in each cell. Text color reflects the generative distribution in A. (C) Synthetic image of clonal marker fluorescence when $\sigma_\alpha = 0.25$. Each nucleus is shaded in accordance with its sampled fluorescence intensity. (D-F) Left to right, increasing the fluorescence ambiguity parameter broadens the overlap in fluorescence levels across gene dosages. (D) Distributions used to generate clonal marker fluorescence levels. Red shading denotes gene dosage. (E) Evenly weighted sum of the generative distributions. (F) Example images of clonal marker fluorescence.

S8 Fig Fraction of nuclei correctly labeled during synthetic benchmarking. Each pixel reflects the average across 50 replicates. Clone size reflects the mean number of cells per clone. Performance improves with increasing clone size and worsens with increasing fluorescence ambiguity.

S9 Fig Spatial context is most informative for large clones with ambiguous fluorescence. (A) MAE of labels assigned using a marginal classifier that neglects spatial context. Performance worsens with increasing fluorescence ambiguity but does not depend upon clone size. (B) Annotation performance relative to the marginal classifier. Color scale reflects the \log_2 fold-change in MAE when spatial context is neglected. Blue indicates that spatial context improves performance.

Acknowledgments

We thank the Howard Hughes Medical Institute for funding Nicolás Peláez through the International Student Fellows program and the Hanna H. Gray Fellowship.

References

1. Oates AC, Gorfinkiel N, González-Gaitán M, Heisenberg CP. Quantitative approaches in developmental biology; 2009. Available from: <http://www.nature.com/articles/nrg2548>.
2. Muzzey D, van Oudenaarden A. Quantitative Time-Lapse Fluorescence Microscopy in Single Cells. *Annu Rev Cell Dev Biol.* 2009;25(1):301–327. doi:10.1146/annurev.cellbio.042308.113408.
3. Stelzer EHK. Light-sheet fluorescence microscopy for quantitative biology. *Nat Methods.* 2014;12(1):23–26. doi:10.1038/nmeth.3219.
4. Truong TV, Supatto W. Toward high-content/high-throughput imaging and analysis of embryonic morphogenesis. *Genesis.* 2011;49(7):555–569. doi:10.1002/dvg.20760.
5. Aghaeepour N, Finak G, Hoos H, Mosmann TR, Brinkman R, Gottardo R, et al. Critical assessment of automated flow cytometry data analysis techniques. *Nat Methods.* 2013;10(3):228–238. doi:10.1038/NMETH.2365.
6. Chen X, Hasan M, Libri V, Urrutia A, Beitz B, Rouilly V, et al. Automated flow cytometric analysis across large numbers of samples and cell types. *Clin Immunol.* 2015;157(2):249–260. doi:10.1016/j.clim.2014.12.009.
7. Pyne S, Maier LM, Lin TI, Wang K, Rossin E, Hu X, et al. Automated high-dimensional flow cytometric data analysis. *Proc Natl Acad Sci.* 2009;106(21):8519–8524. doi:10.1073/pnas.0903028106.
8. Bernstein BE, Brown M, Johnson DS, Liu XS, Nussbaum C, Myers RM, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137. doi:10.1186/gb-2008-9-9-r137.
9. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* 2007;8(2). doi:10.1186/gb-2007-8-2-r19.
10. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9. doi:10.1038/nmeth.1923.
11. Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25(9):1105–1111. doi:10.1093/bioinformatics/btp120.
12. Costes SV, Daelemans D, Cho EH, Dobbin Z, Pavlakis G, Lockett S. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys J.* 2004;86(6):3993–4003. doi:10.1529/biophysj.103.038422.
13. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10(6):845–858. doi:10.1038/nprot.2015.053.
14. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* 2006;7(10):R100. doi:10.1186/gb-2006-7-10-r100.

15. Paintdakhi A, Parry B, Campos M, Irnov I, Elf J, Surovtsev I, et al. Oufiti: An integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol Microbiol.* 2016;99(4):767–777. doi:10.1111/mmi.13264.
16. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: An open-source platform for biological-image analysis; 2012. Available from: <http://www.nature.com/articles/nmeth.2019>.
17. Sommer C, Straehle C, Kothe U, Hamprecht FA. Ilastik: Interactive learning and segmentation toolkit. In: *Proc. - Int. Symp. Biomed. Imaging. IEEE*; 2011. p. 230–233. Available from: <http://ieeexplore.ieee.org/document/5872394/>.
18. Jug F, Pietzsch T, Preibisch S, Tomancak P. Bioimage Informatics in the context of Drosophila research. *Methods.* 2014;68(1):60–73. doi:10.1016/j.ymeth.2014.04.004.
19. Sbalzarini IF. Seeing is believing: Quantifying is convincing: Computational image analysis in biology. *Adv Anatomy, Embryol Cell Biol.* 2016;219:1–39.
20. Schindelin J, Rueden CT, Hiner MC, Eliceiri KW. The ImageJ ecosystem: An open platform for biomedical image analysis; 2015. Available from: <http://doi.wiley.com/10.1002/mrd.22489>.
21. Simpson IT, Price DJ. Pax6; a pleiotropic player in development; 2002. Available from: <http://doi.wiley.com/10.1002/bies.10174>.
22. Parody TR, Muskavitch MAT. The pleiotropic function of Delta during postembryonic development of Drosophila melanogaster. *Genetics.* 1993;135(2):527–539.
23. Shilo BZ, Raz E. Developmental control by the Drosophila EGF receptor homolog DER; 1991. Available from: <https://www.sciencedirect.com/science/article/pii/016895259190261N>.
24. Xu T, Rubin GM. Analysis of genetic mosaics in developing and adult Drosophila tissues. *Development.* 1993;117(4):1223–37. doi:8404527.
25. Xu T, Rubin GM. The effort to make mosaic analysis a household tool. *Development.* 2012;139(24):4501–4503. doi:10.1242/dev.085183.
26. Newsome TP, Asling B, Dickson BJ. Analysis of Drosophila photoreceptor axon guidance in eye-specific mosaics. *Development.* 2000;127(4):851–60.
27. Theodosiou NA, Xu T. Use of FLP/FRT system to study Drosophila development. *Methods A Companion to Methods Enzymol.* 1998;14(4):355–365. doi:10.1006/meth.1998.0591.
28. Struhl G, Basler K. Organizing activity of wingless protein in Drosophila. *Cell.* 1993;doi:10.1016/0092-8674(93)90072-X.
29. Halfar K, Rommel C, Stocker H, Hafen E. Ras controls growth, survival and differentiation in the Drosophila eye by different thresholds of MAP kinase activity. *Development.* 2001;128(9):1687–96.
30. Tomlinson A, Struhl G. Delta/Notch and Boss/Sevenless signals act combinatorially to specify the Drosophila R7 photoreceptor. *Mol Cell.* 2001;7(3):487–95.

31. Yang L, Baker NE. Role of the EGFR/Ras/Raf pathway in specification of photoreceptor cells in the *Drosophila* retina. *Development*. 2001;128(7):1183–91.
32. Huang J, Wu S, Barrera J, Matthews K, Pan D. The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the *Drosophila* homolog of YAP. *Cell*. 2005;122(3):421–434. doi:10.1016/j.cell.2005.06.007.
33. Thompson BJ, Cohen SM. The Hippo Pathway Regulates the bantam microRNA to Control Cell Proliferation and Apoptosis in *Drosophila*. *Cell*. 2006;126(4):767–774. doi:10.1016/j.cell.2006.07.013.
34. Atkins M. *Drosophila* Genetics: The Power of Genetic Mosaic Approaches. In: *Methods Mol. Biol.* vol. 1893. Humana Press, New York, NY; 2019. p. 27–42. Available from: http://link.springer.com/10.1007/978-1-4939-8910-2_{_}2.
35. Enomoto M, Siow C, Igaki T. *Drosophila* as a cancer model. In: *Adv. Exp. Med. Biol.* vol. 1076. Springer, Singapore; 2018. p. 173–194. Available from: http://link.springer.com/10.1007/978-981-13-0529-0_{_}10.
36. Germani F, Bergantinos C, Johnston LA. Mosaic analysis in *Drosophila*. *Genetics*. 2018;208(2):473–490. doi:10.1534/genetics.117.300256.
37. Dai W, Peterson A, Kenney T, Burrous H, Montell DJ. Quantitative microscopy of the *Drosophila* ovary shows multiple niche signals specify progenitor cell fate. *Nat Commun*. 2017;8(1):1244. doi:10.1038/s41467-017-01322-9.
38. Bernasek SM, Lachance JFB, Peláez N, Bakker R, Navarro HT, Amaral LAN, et al. Ratio-based sensing of two transcription factors regulates the transit to differentiation. *bioRxiv*. 2018; p. 430744. doi:<http://dx.doi.org/10.1101/430744>.
39. Ghiglione C, Jouandin P, Cérézo D, Noselli S. The *Drosophila* insulin pathway controls Profilin expression and dynamic actin-rich protrusions during collective cell migration. *Development*. 2018;145(14):dev161117. doi:10.1242/dev.161117.
40. Li K, Baker NE. Regulation of the *Drosophila* ID protein Extra macrochaetae by proneural dimerization partners. *Elife*. 2018;7. doi:10.7554/elife.33967.
41. Bacia K, Petrášek Z, Schwille P. Correcting for spectral cross-talk in dual-color fluorescence cross-correlation spectroscopy. *ChemPhysChem*. 2012;13(5):1221–1231. doi:10.1002/cphc.201100801.
42. Elangovan M, Wallrabe H, Chen Y, Day RN, Barroso M, Periasamy A. Characterization of one- and two-photon excitation fluorescence resonance energy transfer microscopy. *Methods*. 2003;29(1):58–73. doi:10.1016/S1046-2023(02)00283-9.
43. Mort RL. Quantitative analysis of patch patterns in mosaic tissues with ClonalTools software. *J Anat*. 2009;215(6):698–704. doi:10.1111/j.1469-7580.2009.01150.x.
44. Helmuth JA, Paul G, Sbalzarini IF. Beyond co-localization: Inferring spatial interactions between sub-cellular structures from microscopy images. *BMC Bioinformatics*. 2010;11(1):372. doi:10.1186/1471-2105-11-372.

45. Shivanandan A, Radenovic A, Sbalzarini IF. MosaicIA: An ImageJ/Fiji plugin for spatial pattern and interaction analysis. *BMC Bioinformatics*. 2013;14(1):349. doi:10.1186/1471-2105-14-349.
46. Furusawa C, Suzuki T, Kashiwagi A, Yomo T, Kaneko K. Ubiquity of Log-normal Distributions in Intra-cellular Reaction Dynamics. *Biophysics (Oxf)*. 2005;1:25–31. doi:10.2142/biophysics.1.25.
47. Beal J. Biochemical complexity drives log-normal variation in genetic expression. *Eng Biol*. 2017;1(1):55–60. doi:10.1049/enb.2017.0004.
48. Teicher H. Identifiability of Finite Mixtures. *Ann Math Stat*. 1963;34(4):1265–1269. doi:10.1214/aoms/1177703862.
49. Rosvall M, Axelsson D, Bergstrom CT. The map equation. *Eur Phys J Spec Top*. 2009;doi:10.1140/epjst/e2010-01179-1.
50. Katz L. A new status index derived from sociometric analysis. *Psychometrika*. 1953;doi:10.1007/BF02289026.
51. Kamada T, Kawai S. An algorithm for drawing general undirected graphs. *Inf Process Lett*. 1989;31(1):7–15. doi:10.1016/0020-0190(89)90102-6.
52. van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, et al. scikit-image: image processing in Python. *PeerJ*. 2014;doi:10.7717/peerj.453.
53. Nobuyuki Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans Syst Man Cybern*. 1979;.
54. Bugarski M, Mansouri M, Niemann A, Rizk A, Berger P, Ziegler U, et al. Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squash. *Nat Protoc*. 2014;9(3):586–596. doi:10.1038/nprot.2014.037.
55. Peláez N, Gavaldà-Miralles A, Wang B, Navarro HT, Gudjonson H, Rebay I, et al. Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. *Elife*. 2015;4. doi:10.7554/eLife.08924.
56. Zinchuk V, Zinchuk O, Okada T. Quantitative Colocalization Analysis of Multicolor Confocal Immunofluorescence Microscopy Images: Pushing Pixels to Explore Biological Phenomena. *Acta Histochem Cytochem*. 2007;40(4):101–111. doi:10.1267/ahc.07002.
57. Arsenovic PT, Mayer CR, Conway DE. SensorFRET: A Standardless Approach to Measuring Pixel-based Spectral Bleed-through and FRET Efficiency using Spectral Imaging. *Sci Rep*. 2017;7(1). doi:10.1038/s41598-017-15411-8.
58. Kim D, Curthoys NM, Parent MT, Hess ST. Bleed-through correction for rendering and correlation analysis in multi-colour localization microscopy. *J Opt*. 2013;15(9). doi:10.1088/2040-8978/15/9/094011.
59. McMullen PD, Morimoto RI, Amaral LAN. Physically grounded approach for estimating gene expression from microarray data. *Proc Natl Acad Sci*. 2010;107(31):13690–13695. doi:10.1073/pnas.1000938107.

60. Gaudette L, Japkowicz N. Evaluation methods for ordinal classification. In: Lect. Notes Comput. Sci.. vol. 5549 LNAI. Springer, Berlin, Heidelberg; 2009. p. 207–210. Available from: http://link.springer.com/10.1007/978-3-642-01818-3_{ }25.
61. Nguyen TM, Wu QMJ. Gaussian mixture-model-based spatial neighborhood relationships for pixel labeling problems. *IEEE Trans Syst Man Cybern.* 2012;42(1):193–202. doi:10.1109/TSMCB.2011.2161284.
62. Gambis A, Dourlen P, Steller H, Mollereau B. Two-color in vivo imaging of photoreceptor apoptosis and development in *Drosophila*. *Dev Biol.* 2011;351(1):128–134. doi:10.1016/j.ydbio.2010.12.040.
63. Dourlen P, Levet C, Mejat A, Gambis A, Mollereau B. The Tomato/GFP-FLP/FRT Method for Live Imaging of Mosaic Adult *Drosophila* Photoreceptor Cells. *J Vis Exp.* 2013;79. doi:10.3791/50610.
64. Fisher YE, Yang HH, Isaacman-Beck J, Xie M, Gohl DM, Clandinin TR. FlpStop, a tool for conditional gene control in *Drosophila*. *Elife.* 2017;6. doi:10.7554/eLife.22279.
65. Wu JS, Luo L. A protocol for mosaic analysis with a repressible cell marker (MARCM) in *Drosophila*. *Nat Protoc.* 2007;1(6):2583–2589. doi:10.1038/nprot.2006.320.
66. Zhou Q, Neal SJ, Pignoni F. Mutant analysis by rescue gene excision: New tools for mosaic studies in *Drosophila*. *Genesis.* 2016;54(11):589–592. doi:10.1002/dvg.22984.
67. Heffern E, Perrimon N, Hohl AM, del Valle Rodriguez A, Bakal C, Bonvin M, et al. The twin spot generator for differential *Drosophila* lineage analysis. *Nat Methods.* 2009;6(8):600–602. doi:10.1038/nmeth.1349.
68. Yu HH, Kao CF, He Y, Ding P, Kao JC, Lee T. A complete developmental sequence of a *Drosophila* neuronal lineage as revealed by twin-spot MARCM. *PLoS Biol.* 2010;8(8):39–40. doi:10.1371/journal.pbio.1000461.
69. Denes AS, Caussinus E, Affolter M, Kanca O, Percival-Smith A. Raeppli: a whole-tissue labeling tool for live imaging of *Drosophila* development. *Development.* 2013;141(2):472–480. doi:10.1242/dev.102913.
70. Hadjieconomou D, Rotkopf S, Alexandre C, Bell DM, Dickson BJ, Salecker I. Flybow: Genetic multicolor cell labeling for neural circuit analysis in *Drosophila melanogaster*. *Nat Methods.* 2011;8(3):260–266. doi:10.1038/nmeth.1567.
71. Hampel S, Chung P, McKellar CE, Hall D, Looger LL, Simpson JH. *Drosophila* Brainbow: a recombinase-based fluorescence labeling technique to subdivide neural expression patterns. *Nat Methods.* 2011;8(3):253–259. doi:10.1038/nmeth.1566.
72. Neufeld TP, De La Cruz AFA, Johnston LA, Edgar BA. Coordination of growth and cell division in the *Drosophila* wing. *Cell.* 1998;93(7):1183–1193. doi:10.1016/S0092-8674(00)81462-2.
73. Tworoger M, Larkin MK, Bryant Z, Ruohola-Baker H. Mosaic analysis in the *Drosophila* ovary reveals a common Hedgehog- inducible precursor stage for stalk and polar cells. *Genetics.* 1999;

74. Collins RT, Linker C, Lewis J. MAZe: A tool for mosaic analysis of gene function in zebrafish. *Nat Methods*. 2010;7(3):219–223. doi:10.1038/nmeth.1423.
75. Muñoz-Jiménez C, Ayuso C, Dobrzynska A, Torres-Mendéz A, Ruiz PdlC, Askjaer P. An efficient FLP-based toolkit for spatiotemporal control of gene expression in *Caenorhabditis elegans*. *Genetics*. 2017;206(4):1763–1778. doi:10.1534/genetics.117.201012.
76. Wang W, Warren M, Bradley A. Induced mitotic recombination of p53 in vivo. *Proc Natl Acad Sci*. 2007;104(11):4501–4505. doi:10.1073/pnas.0607953104.
77. Meijering E. Cell Segmentation: 50 Years Down the Road. *IEEE Signal Process Mag*. 2012;29(5):140–145. doi:10.1109/msp.2012.2204190.