**Automated analysis of mosaiceye imaginal discs in *Drosophila***

**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 gene 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 walls or cell nuclei and annotate the clones before their data are suitable for analysis. Here, we introduce a computational framework that automates each of these tasks for confocal microscopy images of *Drosophila* eye 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 ecosystem, we aim to contribute to the current move toward automated quantitative analysis among developmental biologists.

**INTRODUCTION**

Quantification will be essential as biologists study increasingly complex facets of organismal development (Oates et al., 2009). 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 (Muzzey and van Oudenaarden, 2009; Stelzer, 2015; Truong and Supatto, 2011), 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 (Aghaeepour et al., 2013; Chen et al., 2015; Pyne et al., 2009), genomics and transcriptomics (Bernstein et al., 2008; Hellemans et al., 2007; Langmead and Salzberg, 2012; Trapnell et al., 2009), and other subdisciplines of biology (Costes et al., 2004; Kelley et al., 2015). Image analysis has proven particularly amenable to automation, with several computer vision tools having gained traction among biologists (Carpenter et al., 2006; Paintdakhi et al., 2016; Schindelin et al., 2012; Sommer et al., 2011). These platforms are popular because they increase productivity, improve the consistency and sensitivity of measurements, and obviate the need for specialized computational proficiency (Jug et al., 2014; Sbalzarini, 2016; Schindelin et al., 2015). 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 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 (Ian Simpson and Price, 2002; Parody and Muskavitch, 1993; Shilo and Raz, 1991).

Mosaic analysis addressed this challenge in *Drosophila* by limiting perturbations to a subset of cells within the larval eye (Xu and Rubin, 2012, 1993), a model system with enduring relevance (Beira and Paro, 2016). The technique yields a heterogeneous tissue comprised of genetically distinct patches of cells, or clones. Clone formation may be restricted to specific developmental contexts by using endogenously-activated promoters to drive trans-chromosomal recombination events (Newsome et al., 2000; Theodosiou and Xu, 1998). The timing of these events determines the number and size of the resultant clones (Struhl and Basler, 1993). Cells within each clone are genetically identical. Perturbations are applied by engineering the dosage of a target gene to differ across clones (Fig. 1B), resulting in clones whose cells are either mutant (−/−), heterozygous (+/−), or homozygous (+/+) for the perturbed gene. Labeling these clones with fluorescent markers enables direct comparison of cells subject to control or perturbation conditions (Fig. 1C). Additional reporters may be used to monitor differences in gene expression or morphology across clones (Fig. 1D). Variants of this strategy led to seminal discoveries in neural patterning (Halfar et al., 2001; Tomlinson and Struhl, 2001; Yang and Baker, 2001) and morphogenesis (Huang et al., 2005; Thompson and Cohen, 2006), and remain popular today (Atkins, 2019; Enomoto et al., 2018; Germani et al., 2018).

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. 1E). The former then defines the stratification under which the latter are compared. We call this strategy quantitative mosaic analysis because it replaces subjective visual comparison with a rigorous statistical alternative. Although many recent studies have deployed this approach (Bernasek et al., 2018; Burrous et al., 2017; Ghiglione et al., 2018; Li and Baker, 2018), qualitative visual comparison remains pervasive in the literature.

We suspect the adoption of quantitative mosaic analysis is hindered by an overwhelming demand for specialized computational skills or, in their stead, extensive manual labor. Researchers must first draw or detect boundaries around individual nuclei in a procedure commonly known as *segmentation* (Fig. 1F). Averaging the pixel intensities within each boundary then yields a fluorescence intensity measurement for each reporter in each identified nucleus (Fig. 1G). The measurements should then be corrected to account for any fluorescence bleedthrough between reporter channels (Fig. 1H).Correction often requires single-reporter calibration experiments to quantify the crosstalk, followed by complex calculations to remedy the data (Bacia et al., 2012; Elangovan et al., 2003). Researchers must then label, or *annotate*, each identified nucleus as mutant, heterozygous, or homozygous for the clonal marker. Annotation is typically achieved through visual inspection (Fig. 1I). Cells carrying zero, one, or two copies of the clonal marker should exhibit low, medium, or high average levels of fluorescence, respectively. However, both measurement and biological noise introduce the possibility that some cells’ measured fluorescence levels may not reliably reflect their genetic identity. Annotation must therefore also consider the spatial context surrounding each nucleus. For instance, a nucleus whose neighbors express high levels of the clonal marker is likely to be homozygous for the clonal marker, even if its individual fluorescence level is comparable to that of heterozygous cells (Fig. 1I, white arrows). With many biological replicates containing thousands of cells each, annotation can quickly become insurmountably tedious. The corrected and labeled measurements are then curated for statistical comparison by excluding those on the border of each clone, and limiting their scope to particular regions of the image field (Fig. 1J). Combined, all of these tasks ultimately burden researchers and raise the barrier for adoption of quantitative analysis strategies.

Automation promises to alleviate this bottleneck, yet the literature bears surprisingly few computational resources designed to support quantitative mosaic analysis. The ClonalTools plugin for ImageJ deploys an image-based approach to measure macroscopic features of clone morphology, but is limited to binary classification of mutant versus non-mutant tissue and offers no functionality for comparing reporter expression across clones (Mort, 2009). Alternatively, the MosaicSuite plugin for ImageJ deploys an array of image processing, segmentation, and analysis capabilities to automatically detect spatial interactions between objects found in separate fluorescence channels (Helmuth et al., 2010; Shivanandan et al., 2013). While useful in many other settings, neither of these tools support automated labeling of individual cells or explicit comparison of clones with single-cell resolution. Most modern studies employing a quantitative mosaic analysis instead report using some form of ad hoc semi-automated pipeline built upon ImageJ (Burrous et al., 2017; Ghiglione et al., 2018; Li and Baker, 2018). We are therefore unaware of any platforms that offer comprehensive support for an automated quantitative mosaic analysis workflow.

We previously published a framework for automated segmentation of cell nuclei in the larval eye (Peláez et al., 2015), in addition to a collection of tools for analyzing reporter expression in the segmented nuclei (Bernasek et al., 2018). Here, we leverage this experience to create an open-source framework for automated quantitative mosaic analysis of eye imaginal discs. The framework supports segmentation, bleedthrough correction, and annotation of confocal microscopy data (Fig. 1F-J). We demonstrate each of these functions by applying them to real confocal images of clones in the eye (Fig. S1A,B), 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.

**Image segmentation**

We implemented a segmentation strategy based upon a standard watershed approach (van der Walt et al., 2014). Briefly, we construct a foreground mask by Otsu thresholding the nuclear stain image following a series of smoothing and contrast-limited adaptive histogram equalization operations (Nobuyuki Otsu, 1979; van der Walt et al., 2014). 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 (Fig. S1C).

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 (Bugarski et al., 2014), including FlyEye Silhouette (Peláez et al., 2015), 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.

**Bleedthrough correction**

Despite efforts to select non-overlapping reporter bandwidths and excite them sequentially, it is not uncommon for reporters excited at one wavelength to emit some fluorescence in another channel (Fig. 1D, yellow lines) (Bacia et al., 2012; Zinchuk et al., 2007). The end result is a positive correlation, or crosstalk, between the measured fluorescence intensities of two or more reporters. Exogenous correlations are problematic given that the purpose of the experiment is to detect changes in reporter levels with respect to the clonal marker.

In our microscopy data, individual clones were distinguished by their low, medium, or high expression levels of an RFP-tagged clonal marker (Fig. 2A). These images should not have shown any detectable difference in GFP levels across clones because all cells carried an equivalent dosage of the control reporter (Fig. S1A). However, the images visibly suffered from bleedthrough between the RFP and GFP channels (Fig. 2A,B). Bleedthrough was similarly evident when we compared measured GFP levels across clones. Nuclei labeled mutant, heterozygous, or homozygous for the clonal marker had low, medium, and high expression levels of the control reporter, respectively (Fig. 2C, black boxes). The data were therefore ripe for systematic correction.

Spectral bleedthrough correction is common practice in other forms of cross-correlation and co-localization microscopy (Bacia et al., 2012; Zinchuk et al., 2007). These methods typically entail characterizing the extent of crosstalk between fluorophores globally (Arsenovic et al., 2017; Kim et al., 2013), on a pixel-by-pixel basis (Elangovan et al., 2003), or by experimental calibration (Bacia et al., 2012), then detrending all images or measurements prior to subsequent analysis. Our framework adopts the global approach, using the background pixels in each image to infer the extent of fluorescence bleedthrough across spectral channels.

Specifically, we assume the fluorescence intensity *Fi,j* for channel *i* at pixel *j* is a superposition of a background intensity *Bi,j* and some function of the expression level *Ei,j* that we seek to compare across cells (McMullen et al., 2010):

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

We further assume that the background intensity of a channel includes linear contributions from the fluorescence intensity of each of the other channels:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

where *k* is indexed over *K* anticipated sources of bleedthrough. Given estimates for each *{α1, α2,…αK}* and *β* we can then estimate the background intensity of each measurement:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

where the braces denote the average across all pixels within a single nucleus. The corrected signal value is obtained by subtracting the background intensity from the measured fluorescence level:

|  |  |  |
| --- | --- | --- |
|  |  | (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, *{α1, α2,…αK}*. We estimate these parameters by characterizing their impact on background pixels (see Methods). When applied to the microscopy data (Fig. S2), this procedure successfully eliminated any detectable difference in GFP expression across clones (Fig. 2C, red boxes).

**Automated annotation of clones**

Our annotation strategy seeks to label each identified cell as mutant, heterozygous, or homozygous 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, 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 (Fig. S3A). This entails fitting a weighted mixture of three or more bivariate lognormal distributions (components) to a two dimensional set of observations (Fig. S3B,C, see Methods for details). 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 (Fig. S3D). 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 (Fig. S3E). We then cluster the components into three groups on the basis of their mean values (Fig. S3F), effectively mapping each component to one of the three possible gene dosages. The model may be trained using observations derived from a single image, or with a collection of observations derived from multiple images. Once trained, the model is able to predict the conditional probability that an individual observation belongs to one of the model’s components, given its measured expression level.

We then use the learned conditional probabilities to detect entire clones, thus assigning a label to each cell. Rather than using the trained model to classify each observation, we compile a new set of observations by limiting each estimate of spatial context to spatially collocated communities with similar expression behavior (Fig. S4A). We identify these communities by applying a community detection algorithm to an undirected graph connecting adjacent cells (Fig. S4B). Edges in this graph are weighted by the similarity of clonal marker expression between neighbors, resulting in communities with similar expression levels (Fig. S4E, Steps I and II). The graph-based approach increases spatial resolution by limiting the information shared by dissimilar neighbors. Applying the mixture model yields an initial estimate of the probability that an observation belongs to one of the model’s components (Fig. S4E, Step III). We further refine these estimates by allowing the probabilities estimated for each cell to diffuse throughout the graph (Fig. S4E, Step IV). The rate of diffusion between neighbors is determined by the weight of the edge that connects them, with more similar neighbors exerting stronger influence on each other. We then use the diffused probabilities to identify the most probable source component and label each observation (Fig. S4E, Step V). These probabilities also provide a measure of confidence in the assigned labels. We replace any low-confidence labels with alternate labels assigned using a marginal classifier that neglects spatial context (Fig. S4F,G), resulting in a fully labeled image (Fig. S4H).

The algorithm leverages the collective wisdom of neighboring measurements to override spatially isolated fluctuations in clonal marker expression, and thereby enforces consistent annotation within contiguous regions of the image field. The size of these regions depends upon the granularity of estimates for the spatial context surrounding each cell. We used an unsupervised approach to choose an appropriate spatial resolution in a principled manner. In short, the resolution is matched to the approximate length scale over which expression levels remain correlated among cells. Both the training and application stages of our annotation algorithm use this automated approach (Figs. S3D and S4D), thus averting any need for user input.

We sought to validate the performance of the annotation algorithm by assessing its ability to accurately reproduce human-assigned labels. We manually labeled nuclei in each eye imaginal disc as mutant, heterozygous, or homozygous for the clonal marker, then automatically labeled the same cells (Fig. 3A). The two sets of labels showed strong overall agreement (Figs. 3B and S5A). Excluding cells on the border of each clone revealed greater than 97% agreement in seven of the nine annotated images (Fig. S5B). Upon secondary inspection of the sole instance of substantial disagreement (Fig. S5C), we are unable to confidently discern which set of labels are more accurate.

**Synthetic benchmarking of annotation performance**

While it is common practice to use human-labeled data as the gold standard, we contend that validation with manually-labeled data entrains implicit human biases in the selection of performant algorithms. These biases are particularly 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 provides a powerful alternative. 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 (Fig. S6A). Cells in this culture contain zero, one, or two copies of a gene encoding an RFP-tagged clonal marker (Fig. S6B). Our simulation procedure ensures that cells tend to remain proximal to their clonal siblings (Fig. S6C), thus forming synthetic clones with tunable size and spatial heterogeneity (Fig. S6D,E). We generated synthetic measurements by randomly sampling fluorescence levels in a dosage-depend manner (Fig. S7A-C). We varied the similarity of fluorescence levels across clones using an ambiguity parameter, *σα*, that modulates the spread of the distributions used to generate fluorescence levels (Fig. S7D-F). Using this schema as a template, we generated a large synthetic dataset, annotated each set of measurements, and compared the assigned labels with their true values. We used the mean absolute error as a comparison metric because it provides a stable measure of accuracy for multiclass classification problems in which the labels are intrinsically ordered (Gaudette and Japkowicz, 2009). Annotation performance is very strong for all cases in which *σα* ≤ 0.3 (Fig. 4). Unsurprisingly, performance suffers as the difficulty of the classification problem is increased. The same trends are evident when performance is graded strictly on accuracy (Fig. S8A). As cells on the periphery of each clone were not excluded from these analyses, the observed metrics provide a lower bound on the performance that may be anticipated in practice.

Performance improved with increasing clone size. We suspected this was caused by larger clones offering additional spatial context to inform the identify of each cell. We verified our assertion by re-evaluating performance relative to a variant of our annotation algorithm that neglects spatial context (Fig. S4G). As expected, the variant’s performance exhibited no dependence on clone size (Fig. S8B). Comparing the two strategies confirmed that spatial context confers the most benefit when clones are large (Fig. S8C). Inclusion of spatial context also becomes increasingly advantageous as the fluorescence ambiguity is increased, even for smaller clones. Thus, spatial context adds progressively more value as the classification task becomes more difficult.

This observation may be rationalized from a statistical perspective. Each cell is classified by maximizing the probability that the assigned label is correct. We compute these probabilities using the estimated expression level of each cell. Neglecting spatial context, this estimate is limited to a single sample and is therefore highly sensitive to both measurement and biological noise. Incorporating spatial context expands the sample size and thereby reduces the standard error of the estimated fluorescence level. The strategy is thus generally well suited to scenarios in which fluorescence intensities correlate across large clones, and closely parallels computer vision methods that exploit spatial contiguity to segment image features with ill-defined borders (Nguyen and Wu, 2012). Because increased measurement precision comes at the expense of spatial resolution, we expect strong performance when measurements are aggregated across relatively large clones, but failure to detect small, heterogeneous clones. These expectations are consistent with the observed results. They are also conveniently aligned with the anticipated properties of real data, as experiments typically attempt to mitigate edge effects by driving early recombination events to generate large clones.

**DISCUSSION**

We used synthetic data to survey the performance of our annotation strategy across a much broader range of conditions than would have otherwise been possible with manually labeled data. This included conditions well beyond those of practical use. In particular, experiments designed to compare gene expression levels across clones would likely seek to avoid generating small clones with ambiguous clonal marker expression. Synthetic data provided a means to survey these edge cases and establish a lower bound on annotation performance. The strong performance observed across the remaining conditions bolsters our confidence that our annotation strategy is well suited to the images it is likely to encounter.

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 (Dourlen et al., 2013; Gambis et al., 2011), but also frequently opt for binary labeling of mutant versus non-mutant clones (Fisher et al., 2017; Wu and Luo, 2007; Zhou et al., 2016) and dichromic labeling of twin-spots (Heffern et al., 2009; Yu et al., 2010). 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 (Denes et al., 2013; Hadjieconomou et al., 2011; Hampel et al., 2011). Our framework is thus well suited to many different mosaic analysis platforms deployed in the larval eye.

In principle, the framework described here should also be applicable to a wide variety of other tissues (Neufeld et al., 1998; Tworoger et al., 1999) and model organisms (Collins et al., 2010; Muñoz-Jiménez et al., 2017; Wang et al., 2007) 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 (Meijering, 2012). 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.

**METHODS**

**Genetics and microscopy of *Drosophila* eye imaginal discs**

We borrowed an experimental control dataset from an unrelated study of neuronal fate commitment during retinal patterning in *Drosophila* (Bernasek et al., 2018). 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 mitotic clones. The chromosome targeted for recombination was marked with a *Ubi-mRFPnls* transgene, enabling automated detection of subpopulations characterized by distinct levels of mRFP fluorescence. The discs also carried a *pnt-GFP* transgene located on a different chromosome that was not subject to mitotic recombination. Discs were dissected, fixed, and co-stained with a DAPI nuclear marker prior to confocal imaging. Please refer to the original study for additional details regarding genetics and experimental conditions.

During eye development, the PntGFP reporter is predominantly expressed in two narrow stripes of progenitor cells (Bernasek et al., 2018). The first stripe occurs immediately posterior to a wave of developmental signaling that traverses the eye disc. Progenitor cells located in this region are suitable for comparison because they are of approximately equivalent developmental age. We applied our automated analysis framework to a total of nine images of these cells.

**Characterization of fluorescence bleedthrough coefficients**

For each image, we morphologically dilate the foreground until no features remain visible (Fig. S2A). We then extract the background pixels and resample them such that the distribution of pixel intensities is approximately uniform (Fig. S2B). Resampling helps mitigate the skewed distribution of pixel intensities found in the background. We then estimate values for each *{α1, α2,…αK}* and *β* by fitting a generalized linear model to the fluorescence intensities of the resampled pixels (Fig. S2C). Each model is a variant of Equation 3 in which angled braces instead denote averages across all background pixels. We formulate these models with identity link functions under the assumption that residuals are gamma distributed. Their coefficients provide an estimate of the bleedthrough contribution strengths that may then be used to estimate the background fluorescence intensity of each nucleus in the corresponding image (Fig. S2D).

**Statistical comparison of fluorescence levels**

To mitigate edge effects, cells residing on the periphery of each clone were excluded from all comparisons (Fig. S2E). Border cells were identified by using the undirected graph constructed during annotation to find all cells connected to a neighbor within a different clone (Fig. S4B). Comparisons were limited to the region of elevated GFP expression near the morphogenetic furrow (Fig. S2F), then further limited to cells undergoing similar stages of development (Fig. S2G). 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.

**Clone annotation algorithm**

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

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

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

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

where *0 ≤ λn ≤ 1* are the mixing proportions, *θn=(, )* are the mean and variance of the *n*th distribution. This assumption is supported by both empirical observations and theoretical insights (Beal, 2017; Furusawa et al., 2005). 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:

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

where *αm* denotes the overall fraction of cells with *m* copies of the gene encoding the clonal marker. For brevity, we substitute yielding:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

Given a collection of sampled fluorescence levels, *{Xi}i=1…N*, we use expectation maximization to find values of *θk* and *λk* 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 smallest value of the Bayesian Information Criterion (BIC):

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

|  |  |  |
| --- | --- | --- |
|  |  | (12) |

where *N* is the sample size, *ln(K)* is the maximum value of the log-likelihood, the subscript *K* denotes the number of mixture components in the model, and *qK* is the total number of parameters (i.e. *K-1* values of *λk* and *2K* values of *μk* and *σ2k*).

Applying Bayes’ rule to the selected model infers the posterior probabilities that each sample *Xi* belongs to the *k*th component:

|  |  |  |
| --- | --- | --- |
|  |  | (13) |

where *p(Xi|k)* is evaluated using the model’s likelihood function and *p(Xi)* 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.

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*:

|  |  |  |
| --- | --- | --- |
|  |  | (14) |

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

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 (Teicher, 1963), 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 variation within each clone precludes accurate classification of a cell’s genotype solely on the basis of its individual expression level. However, clonal lineages are unlikely to exist in isolation because recombination events are usually 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.

We incorporate spatial context by introducing a second jointly-distributed variable *Yi*:

|  |  |  |
| --- | --- | --- |
|  |  | (15) |

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

|  |  |  |
| --- | --- | --- |
|  |  | (16) |

where *μX* and *σ2X* are the global mean and standard deviation, and angled brackets denote the mean across all pairs of cells separated by distance *δ*. We efficiently implement this procedure by fitting an exponential decay function to the down-sampled moving average of *ψ(δ)* as a function of increasing separation distance.

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

|  |  |  |
| --- | --- | --- |
|  |  | (17) |

where contains the mean and variance of each component given by vectors of length two. This formulation constrains each component’s covariance matrix to be diagonal. The posterior is now:

|  |  |  |
| --- | --- | --- |
|  |  | (18) |

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

|  |  |  |
| --- | --- | --- |
|  |  | (19) |

When neglecting spatial context, we use this expression to classify each sample by applying the mapping *f* to the value of *k* that maximizes *p(k|Xi)*:

|  |  |  |
| --- | --- | --- |
|  |  | (20) |

In all other cases, we deploy a graph-based approach to refine the estimate of *p(k|Xi,Yi)*. This first entails constructing an undirected graph connecting adjacent cells within each image. We obtain the graph’s edges through Delaunay triangulation of the measured cell positions, then exclude distant neighbors by thresholding the edge lengths. Each edge is assigned a weight *wij* reflecting the similarity of clonal marker expression between adjacent cells *i* and *j*:

|  |  |  |
| --- | --- | --- |
|  |  | (21) |

where *Eij* is the absolute log fold-change in measured expression level and angled brackets denote the mean across all edges. We chose an exponential formulation because it yields an approximately uniform distribution of edge weights. We then detect communities within the graph using the Infomap algorithm. The algorithm provides a hierarchical partitioning of nodes into non-overlapping clusters. We aggregate all clusters below a critical level that is again chosen by estimating the spatial correlation decay constant. We then enumerate *p(k|Xi,Yci)* where *Yci* is the spatial context obtained by averaging expression levels among all neighbors in the same community as cell *i*.

We further incorporate spatial context by allowing the posterior probabilities *p(k|Xi,Yci)* to diffuse among adjacent cells. We define the modified posterior probability *(k|Xi,Yci)* through a recursive relation analogous to the Katz centrality, initialized by *p(k|Xi,Yci)*:

|  |  |  |
| --- | --- | --- |
|  |  | (22) |

where *α* is the attenuation factor and *wij* are the edge weights. Expressed in matrix form, the solution for *(k|X,Yc)* is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (23) |

where *W* is the matrix of edge weights *wij* and *I* denotes the identity matrix. We then assign a label to each measurement by applying *f* to the value of *k* that maximizes *(k|Xi,Yci)*:

|  |  |  |
| --- | --- | --- |
|  |  | (24) |

Finally, we assess the total posterior probability of each assigned label, *(mi)*:

|  |  |  |
| --- | --- | --- |
|  |  | (25) |

This measure reflects the overall confidence that *mi* is the appropriate label. Labels whose confidence falls below 80% are replaced by their counterparts estimated using the marginal classifier. This substitution helps preserve classification accuracy in situations where spatial context is not informative, and is particularly useful when the annotated clones are relatively small.

**Measurement curation**

Our framework includes a simple graphical user interface that permits manual curation of which regions of the image field are included in subsequent analyses. This tool allows for the exclusion of certain aspects of the segmented data on a context-dependent basis. For example, when studying *Drosophila* eye imaginal discs it is common to restrict analysis to a narrow column of cells in which the cellular processes of interest are taking place (Fig. S2F). Alternatively, users may exclude regions of the image field marked by extensive tissue deformation or cell death. This functionality is convenient but is not a necessary component of a quantitative clonal analysis workflow.

**Cell growth simulations**

We simulated the two dimensional growth of a cell culture seeded with a single cell. In these simulations, growth proceeds through sequential division of cells (Fig. S6A). Not all cells divide at each timestep because cell division is a stochastic process. Instead, each cell divides probabilistically with a rate controlled by a global growth rate parameter.

Cells are simultaneously subject to mitotic recombination (Fig. S6C). Each cell contains either zero, one, or two copies of a gene encoding an RFP-tagged clonal marker. 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 probabilistically with a frequency defined by a global recombination rate parameter.

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 (Fig. S6D). Early recombination events generally entail larger clones, while shorter recombination periods limit the extent of clone formation (Fig. S6E).

After each round of cell division, all cells are repositioned in order to preserve approximately uniform spatial density (Fig. S6C). 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 clonal lineages remain spatially collocated after repositioning. Cell positions are then updated using a force-directed graph drawing algorithm (Kamada and Kawai, 1989). Alternating cell division and repositioning steps are then repeated until a predefined population size is reached.

We used this platform to generate all synthetic data presented in the manuscript. 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. Except where stated otherwise, all simulations limited recombination to the first sixteen cell division events.

**Generation of synthetic microscopy data**

Cell growth simulations yield a list of spatial coordinates and gene dosages for each nucleus (Fig. S6B). The corresponding fluorescence levels *{x1, x2,… xi=N}* were sampled from a lognormal distribution conditioned upon the corresponding gene dosage (Fig. S7A-C):

|  |  |  |
| --- | --- | --- |
|  |  | (26) |

where the subscript *n* denotes the gene copy number and *θn=(μn, σα2)* are the mean and variance of the corresponding distribution. We define *μn* such that the mean fluorescence level doubles for each additional copy of the gene:

|  |  |  |
| --- | --- | --- |
|  |  | (27) |

We refer to *σα* as the fluorescence ambiguity coefficient because it modulates the similarity of fluorescence levels across gene dosages. Increasing *σα* increases the overlap among 𝒩0, 𝒩1, and 𝒩2 (Fig. S7D,E), and consequently increases the difficulty of the annotation task (Fig. S7F).

**Synthetic benchmarking of automated annotation performance**

We created a synthetic microscopy dataset by varying the timing of recombination events to generate cell cultures spanning a range of sixteen average clone sizes (Fig. S6D, only half are shown). We performed 50 replicate simulations for each condition. The mixture model was independently trained and applied to each replicate. Training a single model on all replicates yields 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.

We then labeled each cell and quantified annotation performance by evaluating the mean absolute difference between the predicted labels and their respective true values. This metric preserves ordinality, meaning egregious misclassifications are penalized more severely than mild ones (Gaudette and Japkowicz, 2009). The entire procedure was repeated for sixteen different fluorescence ambiguity coefficient values.

**Data and software availability**

We have distributed the automated mosaic analysis framework as an open-source python package available at https://github.com/sebastianbernasek/clones. We also intend to incorporate its core features into future version 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.These include a series of Jupyter notebooks that enable complete reproduction of all figures and analysis presented in this manuscript.

**REFERENCES**

Aghaeepour N, Finak G, Hoos H, Mosmann TR, Brinkman R, Gottardo R, Scheuermann RH, Dougall D, Khodabakhshi AH, Mah P, Obermoser G, Spidlen J, Taylor I, Wuensch SA, Bramson J, Eaves C, Weng AP, Fortuno ES, Ho K, Kollmann TR, Rogers W, De Rosa S, Dalai B, Azad A, Pothen A, Brandes A, Bretschneider H, Bruggner R, Finck R, Jia R, Zimmerman N, Linderman M, Dill D, Nolan G, Chan C, Khettabi F El, O’Neill K, Chikina M, Ge Y, Sealfon S, Sugár I, Gupta A, Shooshtari P, Zare H, De Jager PL, Jiang M, Keilwagen J, Maisog JM, Luta G, Barbo AA, Májek P, Vilček J, Manninen T, Huttunen H, Ruusuvuori P, Nykter M, McLachlan GJ, Wang K, Naim I, Sharma G, Nikolic R, Pyne S, Qian Y, Qiu P, Quinn J, Roth A, Meyer P, Stolovitzky G, Saez-Rodriguez J, Norel R, Bhattacharjee M, Biehl M, Bucher P, Bunte K, Di Camillo B, Sambo F, Sanavia T, Trifoglio E, Toffolo G, Slavica Dimitrieva SD, Dreos R, Ambrosini G, Grau J, Grosse I, Posch S, Guex N, Kursa M, Rudnicki W, Liu B, Maienschein-Cline M, Schneider P, Seifert M, Strickert M, Vilar JMG. 2013. Critical assessment of automated flow cytometry data analysis techniques. *Nat Methods* **10**:228–238. doi:10.1038/NMETH.2365

Arsenovic PT, Mayer CR, Conway DE. 2017. SensorFRET: A Standardless Approach to Measuring Pixel-based Spectral Bleed-through and FRET Efficiency using Spectral Imaging. *Sci Rep* **7**:15609. doi:10.1038/s41598-017-15411-8

Atkins M. 2019. Drosophila Genetics: The Power of Genetic Mosaic ApproachesMethods in Molecular Biology. Humana Press, New York, NY. pp. 27–42. doi:10.1007/978-1-4939-8910-2\_2

Bacia K, Petrášek Z, Schwille P. 2012. Correcting for spectral cross-talk in dual-color fluorescence cross-correlation spectroscopy. *ChemPhysChem* **13**:1221–1231. doi:10.1002/cphc.201100801

Beal J. 2017. Biochemical complexity drives log-normal variation in genetic expression. *Eng Biol* **1**:55–60. doi:10.1049/enb.2017.0004

Beira J V., Paro R. 2016. The legacy of Drosophila imaginal discs. *Chromosoma*. doi:10.1007/s00412-016-0595-4

Bernasek SM, Lachance J-FB, Peláez N, Bakker R, Navarro HT, Amaral LAN, Bagheri N, Rebay I, Carthew RW. 2018. Ratio-based sensing of two transcription factors regulates the transit to differentiation. *bioRxiv* 430744. doi:http://dx.doi.org/10.1101/430744

Bernstein BE, Brown M, Johnson DS, Liu XS, Nussbaum C, Myers RM, Liu T, Li W, Meyer CA, Eeckhoute J, Zhang Y. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**:R137. doi:10.1186/gb-2008-9-9-r137

Bugarski M, Mansouri M, Niemann A, Rizk A, Berger P, Ziegler U, Sbalzarini IF, Incardona P, Paul G. 2014. Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squassh. *Nat Protoc* **9**:586–596. doi:10.1038/nprot.2014.037

Burrous H, Kenney T, Dai W, Montell DJ, Peterson A. 2017. Quantitative microscopy of the Drosophila ovary shows multiple niche signals specify progenitor cell fate. *Nat Commun* **8**:1244. doi:10.1038/s41467-017-01322-9

Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, Golland P, Sabatini DM. 2006. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7**:R100. doi:10.1186/gb-2006-7-10-r100

Chen X, Hasan M, Libri V, Urrutia A, Beitz B, Rouilly V, Duffy D, Patin É, Chalmond B, Rogge L, Quintana-Murci L, Albert ML, Schwikowski B. 2015. Automated flow cytometric analysis across large numbers of samples and cell types. *Clin Immunol* **157**:249–260. doi:10.1016/j.clim.2014.12.009

Collins RT, Linker C, Lewis J. 2010. MAZe: A tool for mosaic analysis of gene function in zebrafish. *Nat Methods* **7**:219–223. doi:10.1038/nmeth.1423

Costes S V., Daelemans D, Cho EH, Dobbin Z, Pavlakis G, Lockett S. 2004. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys J* **86**:3993–4003. doi:10.1529/biophysj.103.038422

Denes AS, Caussinus E, Affolter M, Kanca O, Percival-Smith A. 2013. Raeppli: a whole-tissue labeling tool for live imaging of Drosophila development. *Development* **141**:472–480. doi:10.1242/dev.102913

Dourlen P, Levet C, Mejat A, Gambis A, Mollereau B. 2013. The Tomato/GFP-FLP/FRT Method for Live Imaging of Mosaic Adult Drosophila Photoreceptor Cells. *J Vis Exp* e50610. doi:10.3791/50610

Elangovan M, Wallrabe H, Chen Y, Day RN, Barroso M, Periasamy A. 2003. Characterization of one- and two-photon excitation fluorescence resonance energy transfer microscopy. *Methods* **29**:58–73. doi:10.1016/S1046-2023(02)00283-9

Enomoto M, Siow C, Igaki T. 2018. Drosophila as a cancer modelAdvances in Experimental Medicine and Biology. Springer, Singapore. pp. 173–194. doi:10.1007/978-981-13-0529-0\_10

Fisher YE, Yang HH, Isaacman-Beck J, Xie M, Gohl DM, Clandinin TR. 2017. FlpStop, a tool for conditional gene control in Drosophila. *Elife* **6**. doi:10.7554/eLife.22279

Furusawa C, Suzuki T, Kashiwagi A, Yomo T, Kaneko K. 2005. Ubiquity of Log-normal Distributions in Intra-cellular Reaction Dynamic. *Biophys (Nagoya-shi, Japan)* **1**:25–31. doi:10.2142/biophysics.1.25

Gambis A, Dourlen P, Steller H, Mollereau B. 2011. Two-color in vivo imaging of photoreceptor apoptosis and development in Drosophila. *Dev Biol* **351**:128–134. doi:10.1016/j.ydbio.2010.12.040

Gaudette L, Japkowicz N. 2009. Evaluation methods for ordinal classificationLecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Berlin, Heidelberg. pp. 207–210. doi:10.1007/978-3-642-01818-3\_25

Germani F, Bergantinos C, Johnston LA. 2018. Mosaic analysis in drosophila. *Genetics* **208**:473–490. doi:10.1534/genetics.117.300256

Ghiglione C, Jouandin P, Cérézo D, Noselli S. 2018. The Drosophila insulin pathway controls Profilin expression and dynamic actin-rich protrusions during collective cell migration. *Development* **145**:dev161117. doi:10.1242/dev.161117

Hadjieconomou D, Rotkopf S, Alexandre C, Bell DM, Dickson BJ, Salecker I. 2011. Flybow: Genetic multicolor cell labeling for neural circuit analysis in Drosophila melanogaster. *Nat Methods* **8**:260–266. doi:10.1038/nmeth.1567

Halfar K, Rommel C, Stocker H, Hafen E. 2001. Ras controls growth, survival and differentiation in the Drosophila eye by different thresholds of MAP kinase activity. *Development* **128**:1687–96.

Hampel S, Chung P, McKellar CE, Hall D, Looger LL, Simpson JH. 2011. Drosophila Brainbow: a recombinase-based fluorescence labeling technique to subdivide neural expression patterns. *Nat Methods* **8**:253–259. doi:10.1038/nmeth.1566

Heffern E, Perrimon N, Hohl AM, del Valle Rodriguez A, Bakal C, Bonvin M, Sustar A, Grunwald D, Griffin R, Villalta C, Wu C, Bateman JR, Desplan C, Binari R, Schubiger G. 2009. The twin spot generator for differential Drosophila lineage analysis. *Nat Methods* **6**:600–602. doi:10.1038/nmeth.1349

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. 2007. qBaserelativequantificationframeworkandsoftwareformanagementandautomatedanalysisofreal-timequantitativePCRdata. *Genomebiology* **8**:R19. doi:10.1186/gb-2007-8-2-r19

Helmuth JA, Paul G, Sbalzarini IF. 2010. Beyond co-localization: Inferring spatial interactions between sub-cellular structures from microscopy images. *BMC Bioinformatics* **11**:372. doi:10.1186/1471-2105-11-372

Huang J, Wu S, Barrera J, Matthews K, Pan D. 2005. The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the Drosophila homolog of YAP. *Cell* **122**:421–434. doi:10.1016/j.cell.2005.06.007

Ian Simpson T, Price DJ. 2002. Pax6; a pleiotropic player in development. *BioEssays*. doi:10.1002/bies.10174

Jug F, Pietzsch T, Preibisch S, Tomancak P. 2014. Bioimage Informatics in the context of Drosophila research. *Methods* **68**:60–73. doi:10.1016/j.ymeth.2014.04.004

Kamada T, Kawai S. 1989. An algorithm for drawing general undirected graphs. *Inf Process Lett* **31**:7–15. doi:10.1016/0020-0190(89)90102-6

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**:845–858. doi:10.1038/nprot.2015.053

Kim D, Curthoys NM, Parent MT, Hess ST. 2013. Bleed-through correction for rendering and correlation analysis in multi-colour localization microscopy. *J Opt (United Kingdom)* **15**. doi:10.1088/2040-8978/15/9/094011

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–9. doi:10.1038/nmeth.1923

Li K, Baker NE. 2018. Regulation of the Drosophila ID protein Extra macrochaetae by proneural dimerization partners. *Elife* **7**. doi:10.7554/elife.33967

McMullen PD, Morimoto RI, Amaral LAN. 2010. Physically grounded approach for estimating gene expression from microarray data. *Proc Natl Acad Sci* **107**:13690–13695. doi:10.1073/pnas.1000938107

Meijering E. 2012. Cell Segmentation: 50 Years Down the Road. *IEEE Signal Process Mag* **29**:140–145. doi:10.1109/msp.2012.2204190

Mort RL. 2009. Quantitative analysis of patch patterns in mosaic tissues with ClonalTools software. *J Anat* **215**:698–704. doi:10.1111/j.1469-7580.2009.01150.x

Muñoz-Jiménez C, Ayuso C, Dobrzynska A, Torres-Mendéz A, Ruiz P de la C, Askjaer P. 2017. An efficient flp-based toolkit for spatiotemporal control of gene expression in caenorhabditis elegans. *Genetics* **206**:1763–1778. doi:10.1534/genetics.117.201012

Muzzey D, van Oudenaarden A. 2009. Quantitative Time-Lapse Fluorescence Microscopy in Single Cells. *Annu Rev Cell Dev Biol* **25**:301–327. doi:10.1146/annurev.cellbio.042308.113408

Neufeld TP, De La Cruz AFA, Johnston LA, Edgar BA. 1998. Coordination of growth and cell division in the Drosophila wing. *Cell* **93**:1183–1193. doi:10.1016/S0092-8674(00)81462-2

Newsome TP, Asling B, Dickson BJ. 2000. Analysis of Drosophila photoreceptor axon guidance in eye-specific mosaics. *Development* **127**:851–60.

Nguyen TM, Wu QMJ. 2012. Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem. *IEEE Trans Syst Man, Cybern Part B Cybern* **42**:193–202. doi:10.1109/TSMCB.2011.2161284

Nobuyuki Otsu. 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans SYSTREMS, MAN, Cybern*.

Oates AC, Gorfinkiel N, González-Gaitán M, Heisenberg C-P. 2009. Quantitative approaches in developmental biology. *Nat Rev Genet* **10**:517–530. doi:10.1038/nrg2548

Paintdakhi A, Parry B, Campos M, Irnov I, Elf J, Surovtsev I, Jacobs-Wagner C. 2016. Oufti: An integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol Microbiol* **99**:767–777. doi:10.1111/mmi.13264

Parody TR, Muskavitch MAT. 1993. The pleiotropic function of Delta during postembryonic development of Drosophila melanogaster. *Genetics* **135**:527–539.

Peláez N, Gavalda-Miralles A, Wang B, Navarro HT, Gudjonson H, Rebay I, Dinner AR, Katsaggelos AK, Amaral LAN, Carthew RW. 2015. Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. *Elife* **4**. doi:10.7554/eLife.08924

Pyne S, Maier LM, Lin T-I, Wang K, Rossin E, Hu X, Tamayo P, Hafler DA, De Jager PL, McLachlan GJ, Baecher-Allan C, Mesirov JP. 2009. Automated high-dimensional flow cytometric data analysis. *Proc Natl Acad Sci* **106**:8519–8524. doi:10.1073/pnas.0903028106

Sbalzarini IF. 2016. Seeing is believing: Quantifying is convincing: Computational image analysis in biology. *Adv Anat Embryol Cell Biol* **219**:1–39. doi:10.1007/978-3-319-28549-8\_1

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: An open-source platform for biological-image analysis. *Nat Methods*. doi:10.1038/nmeth.2019

Schindelin J, Rueden CT, Hiner MC, Eliceiri KW. 2015. The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol Reprod Dev*. doi:10.1002/mrd.22489

Shilo BZ, Raz E. 1991. Developmental control by the Drosophila EGF receptor homolog DER. *Trends Genet*. doi:10.1016/0168-9525(91)90261-N

Shivanandan A, Radenovic A, Sbalzarini IF. 2013. MosaicIA: An ImageJ/Fiji plugin for spatial pattern and interaction analysis. *BMC Bioinformatics* **14**:349. doi:10.1186/1471-2105-14-349

Sommer C, Straehle C, Kothe U, Hamprecht FA. 2011. Ilastik: Interactive learning and segmentation toolkitProceedings - International Symposium on Biomedical Imaging. IEEE. pp. 230–233. doi:10.1109/ISBI.2011.5872394

Stelzer EHK. 2015. Light-sheet fluorescence microscopy for quantitative biology. *Nat Methods* **12**:23–26. doi:10.1038/nmeth.3219

Struhl G, Basler K. 1993. Organizing activity of wingless protein in Drosophila. *Cell*. doi:10.1016/0092-8674(93)90072-X

Teicher H. 1963. Identifiability of Finite Mixtures. *Ann Math Stat* **34**:1265–1269. doi:10.1214/aoms/1177703862

Theodosiou NA, Xu T. 1998. Use of FLP/FRT system to study Drosophila development. *Methods A Companion to Methods Enzymol* **14**:355–365. doi:10.1006/meth.1998.0591

Thompson BJ, Cohen SM. 2006. The Hippo Pathway Regulates the bantam microRNA to Control Cell Proliferation and Apoptosis in Drosophila. *Cell* **126**:767–774. doi:10.1016/j.cell.2006.07.013

Tomlinson A, Struhl G. 2001. Delta/Notch and Boss/Sevenless signals act combinatorially to specify the Drosophila R7 photoreceptor. *Mol Cell* **7**:487–95.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**:1105–1111. doi:10.1093/bioinformatics/btp120

Truong T V., Supatto W. 2011. Toward high-content/high-throughput imaging and analysis of embryonic morphogenesis. *Genesis* **49**:555–569. doi:10.1002/dvg.20760

Tworoger M, Larkin MK, Bryant Z, Ruohola-Baker H. 1999. Mosaic analysis in the Drosophila ovary reveals a common Hedgehog- inducible precursor stage for stalk and polar cells. *Genetics*.

van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T. 2014. scikit-image: image processing in Python. *PeerJ*. doi:10.7717/peerj.453

Wang W, Warren M, Bradley A. 2007. Induced mitotic recombination of p53 in vivo. *Proc Natl Acad Sci* **104**:4501–4505. doi:10.1073/pnas.0607953104

Wu CFJ. 1983. On the Convergence Properties of the EM Algorithm. *Ann Stat* **11**:95–103. doi:10.1214/aos/1176346060

Wu JS, Luo L. 2007. A protocol for mosaic analysis with a repressible cell marker (MARCM) in Drosophila. *Nat Protoc* **1**:2583–2589. doi:10.1038/nprot.2006.320

Xu T, Rubin GM. 2012. The effort to make mosaic analysis a household tool. *Development* **139**:4501–4503. doi:10.1242/dev.085183

Xu T, Rubin GM. 1993. Analysis of genetic mosaics in developing and adult Drosophila tissues. *Development* **117**:1223–37. doi:8404527

Yang L, Baker NE. 2001. Role of the EGFR/Ras/Raf pathway in specification of photoreceptor cells in the Drosophila retina. *Development* **128**:1183–91.

Yu HH, Kao CF, He Y, Ding P, Kao JC, Lee T. 2010. A complete developmental sequence of a Drosophila neuronal lineage as revealed by twin-spot MARCM. *PLoS Biol* **8**:39–40. doi:10.1371/journal.pbio.1000461

Zhou Q, Neal SJ, Pignoni F. 2016. Mutant analysis by rescue gene excision: New tools for mosaic studies in Drosophila. *Genesis* **54**:589–592. doi:10.1002/dvg.22984

Zinchuk V, Zinchuk O, Okada T. 2007. Quantitative Colocalization Analysis of Multicolor Confocal Immunofluorescence Microscopy Images: Pushing Pixels to Explore Biological Phenomena. *ACTA Histochem Cytochem* **40**:101–111. doi:10.1267/ahc.07002

**FIGURE LEGENDS**

**Figure 1: A framework for conducting quantitative clonal analysis.**

**(A,B)** Experimental framework using mitotic clones to test whether or not regulatory interactions occur between a perturbation target and reporter of interest. Blue and green ovals 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 ovals represent a gene encoding a clonal marker used to identify the resultant clones.

**(C,D)** Conventional clonal analysis in the larval fruit fly eye. (C) Clones are identified by visual comparison of clonal marker fluorescence among nuclei. (D) Regions labeled mutant (−/−) or homozygous (+/+) for the clonal marker are compared with those labeled heterozygous (+/−) to assess whether reporter expression differs across clones. Fluorescence bleed-through is arbitrarily diagnosed.

**(E-J)** Quantitative clonal analysis. Panels depict a magnified view of the region enclosed by red rectangles in panels C and D. (E) Raw confocal image of the nuclear stain, clonal marker, and reporter of interest. (F) Segmentation identifies distinct nuclei. (G) Reporter expression is quantified by averaging the pixel intensities within each segment. Numbers reflect measured values. (H) Measurements may be corrected to mitigate fluorescence bleedthrough. (I) Individual nuclei are labeled mutant, heterozygous, or homozygous for the clonal marker. White arrows mark nuclei with ambiguous fluorescence levels. (J) 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 larval eye, comparison is often limited to a narrow window near the MF (orange arrow).

**Figure 2: 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 Fig. S8E). 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*).

**Figure 3: 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 overlayed on the RFP channel of the image shown in Fig. S1B. 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.

**Figure 4: Synthetic benchmarking of automated annotation performance.**

Grid shows the mean absolute error (MAE) between assigned labels and the corresponding ground truth as a function of fluorescence ambiguity and clone size. Values shown reflect the average across 50 replicate simulations. Performance improves with increasing clone size and worsens with increasing fluorescence ambiguity. Clone size reflects the mean number of cells per clone.

**SUPPLEMENTAL FIGURE LEGENDS**

**Figure S1: 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.

**Figure S2: Intermediate stages of bleedthrough correction.**

**(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.

**Figure S3: 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 Fig. S3C.

**(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 (*ki*). 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 (*mi*) 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, *ψ(δ)*. Black line shows the moving average of *ψ(δ)*, 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.

**Figure S4: Label assignment using a trained clone annotation model.**

**(A)** The measurements *X* from a segmented image are used to sample the spatial context *YC* of the community surrounding each cell before the mixture model is applied (blue and green path). They are simultaneously labeled using a marginal projection of the trained model (magenta path). The two sets of labels are then merged (red path). Subsequent panels demonstrate these procedures using the example shown in Fig. S3C.

**(B-D)** Spatial context sampling. (B) Weighted undirected graph connecting adjacent cells. Line thickness reflects the expression similarity between neighbors.(C) Community resolution is defined by aggregating clusters that fall below a cut level (*δ*) in the hierarchy. Images show potential 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, *ψ(δ)* (black line). In this example, clusters are aggregated below the fifth level. Panel E instead depicts aggregation below the third level for ease of visualization.

**(E)** Application of the mixture model. *(I)* The graph connecting adjacent cells 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 entire graph. *(V)* Each cell is assigned its most probable label.

**(F,G)** Application of a marginal mixture model that neglects spatial context. (F) Marginal mixture model components obtained by summing across the spatial context dimension of the full mixture model. Red shading denotes the label assigned to each component. Dashed black line is the overall marginal density. (G) Marginal classifier that labels cells strictly on the basis of their individual fluorescence level. Red shading denotes the most probable label for each expression level.

**(H)** Annotated measurements. Red shading denotes the assigned label. Labels with low confidence (*(mi) < 0.8*) are replaced by their marginal counterparts.

**Figure S5: Comparison of automated annotation with manually assigned labels.**

**(A)** Distribution of labels among each possible value.

**(B)** Agreement between automated and manual annotation. Values in parentheses exclude cells on the periphery of each clone. The sole instance of low agreement is marked in red.

**(C)** Visual comparison of the instance in which automated and manual annotation differ. Image shows clonal marker fluorescence, overlayed colors denote the assigned label.

**Figure S6: 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 (*Δt=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.

**Figure S7: 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 *σα=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.

**Figure S8: Additional results for synthetic benchmarking of annotation algorithm.**

**(A)** Annotation accuracy as a function of fluorescence ambiguity and clone size. Accuracy is defined as the fraction of cells that were correctly labeled. Performance improves with increasing clone size and worsens with increasing fluorescence ambiguity.

**(B)** Annotation performance of a marginal classifier that neglects spatial context. Each point in the grid reflects the mean absolute error (MAE) between assigned labels and the corresponding ground truth. Performance worsens with increasing fluorescence ambiguity but does not depend upon clone size.

**(C)** Annotation performance relative to the marginal classifier. Color scale reflects the log2 fold-change in MAE when spatial context is ignored. Blue indicates that spatial context improves performance. Spatial context is informative for larger clones, particularly when fluorescence levels are ambiguous.