1   **Accessible and accurate cytometry analysis using fluorescence microscopes**

2   Daniel Foyt[1, *], Yiming Kuang[2], Samma Rehem[3], Klaus Yserentant[2], Bo Huang[2,4,5,*]

3

4   [1] UCSF-UC Berkeley Joint Graduate Program in Bioengineering, University of California San

5   Francisco, San Francisco, California, 94143, United States of America

6   [2] Department of Pharmaceutical Chemistry, University of California San Francisco, San

7   Francisco, California, 94143, United States of America

8   [3] Barnard College, New York City, New York, 10027, United States of America

9   [4] Department of Biochemistry and Biophysics, University of California San Francisco, San

10   Francisco, California, 94143, United States of America

11   [5] Chan Zuckerberg Biohub San Francisco, San Francisco, California, 94158, United States of

12   America

13   [*] To whom correspondence should be addressed. Tel1: +1 4154761866; Fax: +1 4155141028;

14   Email: daniel_foyt@berkeley.edu, bo.huang@ucsf.edu.

15

16

17

18

19

20

21

22

23

24

25

26

27

28

Abstract

30   We have developed a method along with a python-based analysis tool to capture images

31   and produce flow cytometry like data utilizing simple accessible microscopes. Utilizing the

32   recently developed generalist algorithms for cell segmentation, our approach easily segments

33   semi-adherent or suspended cells facilitating quantification of fluorescent intensity similar to flow

34   cytometry. We have shown that our approach exhibits similar speed and enhanced sensitivity

35   when compared to typical flow cytometry. The utility of our approach is demonstrated by

36   screening a set of 88 prime editing conditions utilizing the integration of mNeonGreen$_{11}$ as a

37   reporter.

38

Introduction

40   Measurements of the number, size, physical phenotype, and presence or absence of a

41   specific protein or reporter in a pool of cells are a crucial part of a biologist tool kit. Currently,

42   these functions are typically carried out using flow cytometry. Flow cytometry operates by

43   flowing single cells in a steady aligned stream using fluid flow techniques and analyzing droplets

44   or cells as they pass through a laser or electrical field (1). Flow cytometry has been a very

45   powerful method in the field of immunology and molecular biology to measure the fluorescent

46   readout of thousands of cells (1). While powerful in its throughput, the information obtained from

47   this technique is inherently lacking in depth as it typically only provides an average intensity of

48   the fluorescent reporter while not providing spatial information about the signal distribution in the

49   context of other cellular components. Additionally, the necessity to suspend cells in solution for

50   analysis can be limiting in some cases. Historically, more detailed measurements of cellular

51   physical phenotypes were made with microscopy, which would necessitate compromising the

52   scale and speed of the assay compared to flow cytometry. Moreover, segmentation of images

53   so that individual cells growing in groups can be distinguished is challenging without explicit

54   reporters for cell boundaries(2,3). With the advancement of microscope automation, image

55   analysis tools, and the availability of suitable microscopes to many biologists, microscopy-based

56   cytometry could be come a more accessible alternative to flow cytometry.

57

58   **Results**

59   **Development and validation of microscopy-based cytometry**

60   One of the main impediments to a microscopy-based cytometry method for adherent

61   cells is the ability to segment individual cells reliably when they grow in a cluster. While there

62   are algorithms developed for this task, such as Cellpose(4), a generalist deep learning-based

63   segmentation method, they do not always reliably define cell boundaries in bright-field images

64   due to cells physically overlapping and lack of contrast for cell-cell interfaces(2,3). Although

65   fluorescent staining for cell boundaries (such as plasma membrane stain) can facilitate

66   segmentation, it consumes a color channel of the microscope. Therefore, we sought to develop

67   an approach for reliable segmentation using bright-field images only. We reasoned that inducing

68   cells to partially detach from one another and the surface would reduce their overlap, making

69   the images easier to segment. To determine if partially detached cells facilitate better

70   segmentation, we first imaged cells that were adhered to the surface of a coverglass-bottom 96-

71   well plate using a Nikon Ti-E inverted wide-field microscope and attempted to segment the

72   images using Cellpose(4). As expected, brightfield segmentations match poorly with

73   fluorescence images from nuclear staining, which served as the ground truth control (Fig. 1a).

74   We then treated the cells in the well with trypsin, which is used in flow-cytometry analysis of

75   adherent cells to detach the cells. We observed that, with a brief trypsin treatment, cells stayed

76   attached to the surface but that the boundaries between cells were much more apparent,

77   improving the segmentation results (Fig. 1b). To see if the complete detachment of cells from

78   the surface would further improve the discrimination of cell boundaries, we treated cells with

3

79   trypsin again, mechanically agitated the media to detach the cells (similar to the generation of

80   cell suspension for common flow cytometry preparation procedure), and then centrifuged the

81   well plate to let the cells settle down to the surface for image capture. Segmentation of these

82   images resulted in only a slight improvement in cell segmentation (Fig. 1c). For quantitative

83   assessment, we compared the number of segments obtained from the brightfield images with

84   the number of nuclei counted. Treating cells with trypsin alone and trypsin plus resuspension

85   substantially improved the accuracy compared to adherent conditions (Fig. 1d). Additionally,

86   these treatments decreased the false negative (Fig. 1e) and false positive (Fig. 1f) ratios for cell

87   detection.

88

89   Cellpose utilizes two parameters when determining the boundary of a cell; the first is an

90   estimated segment diameter, as the models used were trained on scaled images with constant

91   diameters, which may be different from the experimental data. The second is a flow threshold,

92   which is the maximum allowed error of the flows determined by the model and the predicted

93   segments. To ensure we utilized the optimal parameters in Cellpose and that the difference in

94   segmentation accuracy was not due to improper parameter selection, we screened a

95   combination of 25 diameters ranging from 25 to 505 pixels and 21 flow threshold values from

96   0.05 to 2.05 for a total of 525 parameter combinations. We calculated the number of segments

97   obtained from the Cellpose segmentation of brightfield images as a fraction of the number of

98   nuclei counted. We observed that for untreated adherent cells, Cellpose only produced 30% of

99   the number of segments as there were nuclei with optimal parameters (cell diameter = 145, flow

100   threshold = 0.95-2.05) (Fig. 2a). The Cellpose segments/nuclei measurements for the

101   trypsinized-only cells had a maximum value of close to 1 with optimal parameters of cell

102   diameter = 145 and flow threshold = 0.95-2.05 (Fig. 2b). The optimal parameters (Cellpose

103   segments/nuclei = 1) for trypsinized and resuspended cells were in a slightly wider range. Still,

104   there were combinations of Cellpose parameters (45 < diameter (pixels) < 265, flow threshold >

105   0.45) that yielded more segments than counted nuclei (Fig. 2c). We also calculated false

106   negative and false positive rates for all parameter combinations and trypsinization conditions.

107   The adhered cell condition had a much higher false negative and false positive ratio compared

108   to the trypsinized and trypsinized and resuspended conditions (Fig. 2d-i).

109

110   To understand the upper limit of the number of cells we could reliably segment and

111   assay in our implementation, we varied the number of cells loaded into a well in a 96-well plate

112   from 15,600 to 500,000, took brightfield images, and segmented the images. Cell densities from

113   15,600 to 62,500 cells/well were reliably segmented (Fig. 3a), and the number of segments

114   linearly correlated with the number of cells loaded. At higher densities, the segmentation

115   efficiency worsens (Fig. 3b). At densities greater than 62,500 cells/well, the cells no longer form

116   a single layer, and all of the cells can not be reliably segmented (Fig. 3a 125k).

117   We performed a cell mixing experiment to determine the single-cell purity of the data obtained

118   from our method. We obtained two cell lines, each expressing spectrally district fluorescent

119   proteins; one HEK293 cell line stably expressing mApple created via landing pad integration (5),

120   and a HeLa cell line with HspA8 endogenously tagged with the mNeonGreen2$_{11}$ fragment (6)

121   (mNG2$_{11}$) and stably expressing mNG3A$_{1-10}$ (7) which yields green fluorescent HspA8 protein.

122   We mixed these cell lines in equal proportions, seeded them into wells, and allowed the cells to

123   adhere and grow. We gently trypsinized these cells and captured images in the brightfield,

124   mNG, and mApple channels (Fig. 4a). We observed that 0.90% of cells contained a mix of

125   signals from both mNG and mApple, both above the wild-type control signal, suggesting a

126   doublet rate of 1.80% (from a 50/50 mix of cells, half of the doublets will arise from two cells with

127   the same fluorescence protein) (Fig. 4b).

128   Next, we compared the specificity of our method to that of flow cytometry. We analyzed HEK

129   cells stably expressing the fluorescent protein mApple and nonfluorescent wild-type (WT) HEK

130   cells. To ensure we are making a fair comparison between flow cytometry and our imaging

131   systems, we optimized the flow cytometer's PMT voltage setting to maximize the signal-to-noise

132   ratio (SNR) and minimize the coefficient of variation (CV) in mApple signal (Fig. 4c-d). We

133   chose a PMT voltage of 450 mV as it had a good SNR, and increasing the voltage did not

134   improve the CV (Fig. 4d). We compared SNR and CV values collected with these optimized flow

135   settings to our approach utilizing two microscopes, including a more accessible, low-cost EVOS

136   microscope. With the same cell population that we measured with flow cytometry, we captured

137   images on our Nikon Ti-E widefield microscope with 20X and 60X magnification and on an

138   EVOS microscope with 10X magnification. The distribution of mApple signal and WT

139   background was comparable on all systems, with the notable exception that the width of the WT

140   distribution for data obtained with the Nikon Ti-E was noticeably less than data obtained with

141   flow or the EVOS microscope (Fig. 4e). We quantified this observation by calculating SNR for all

142   four approaches. All imaging methods had an SNR significantly more than that of optimized flow

143   cytometry (Fig. 4f). We also compared the spread or variation in signals obtained from the

144   mApple HEK cells as we suspected that our imaging-based method might lead to higher levels

145   of variation in fluorescence intensities. Indeed, we observed higher variance in the fluorescence

146   intensities for the data obtained with the EVOS and 20X Nikon Ti-E than with flow cytometry but

147   did not observe a significant difference in CV between 60X Nikon Ti-E data and flow cytometry

148   data (Fig. 4g).

149

150   **Imaging-based cytometry for prime editing optimization**

151

152   Prime editing is a genetic editing technique that utilizes a prime editor (PE), a single

153   strand nicking Cas9 endonuclease fused to an engineered reverse transcriptase enzyme, a

154   prime editing guide RNA (pegRNA) to edit or insert a new sequence into the genome(8), and a

155   traditional single guide RNA (sgRNA) to facilitate opposing strand nicking. Instead of encoding

156   base modification or insertions on exogenously added DNA, the genetic alterations in prime

157   editing are encoded on the pegRNA and are copied into the genome via the reverse

158   transcriptase (Fig. 5a). Prime editing does not make double-stranded breaks in the DNA like

159   traditional CRISPR-Cas9 approaches, which are more error-prone[8]. Along with the ability to

160   edit post-mitotic cells and its potential to address almost 90% of disease-causing mutations,

161   prime editing is an important technology in advancing genetic medicine. While prime editing is

162   reasonably efficient at making small insertions up to 15 bases, the editing efficiency for inserts

163   from 15 bases to 50 bases is inefficient and variable, depending on the genomic context[9]. To

164   improve prime editing, groups have optimized the molecular components of the prime editing

165   system such as the pegRNA sequence[10] and the PE protein structure[11] and studied the

166   endogenous DNA repair pathways involved in prime editing[12]. It was shown that DNA

167   mismatch repair (MMR) pathways strongly suppress the efficiency of prime editing. This

168   information was used to develop a prime editing approach that inhibits MMR by transiently

169   expressing a dominant negative MMR protein (MLH1dn)[12]. Inefficiencies are also partially due

170   to the suboptimal concentrations of the prime editing components (PE, pegRNA, sgRNA,

171   MLH1dn). We reasoned that with the optimization of prime editing conditions, insertions of

172   longer lengths could be made more reliable. One application that would benefit from efficient

173   insertion is integration of the small strand of split fluorescent proteins for endogenous protein

174   tagging. Endogenous protein tagging with split fluorescence proteins functions by genetically

175   appending a small, ~50 bp in most cases, tag to the coding region of a protein of interest. This

176   results in the addition of amino acids of the small non-fluorescence fragment of a fluorescent

177   protein being incorporated into the protein of interest. In the presence of the nonfluorescent

178   large fragment, the small and large fragments combined to form the complete fluorescent form

179   of a reporter[13].

180   To demonstrate the utility of our microscopy cytometry method, we designed a pegRNA and

181   sgRNA targeting the N-terminus of the beta-actin gene encoding an insertion of the mNG2$_{11}$

182   fragment (48 bp). We screened a set of 88 prime editing conditions utilizing positive mNG

7

183   fluorescence as a readout of prime editing insertion of mNG2$_{11}$. We varied the PE plasmid

184   concentrations and the MLH1dn plasmid concentration and assessed editing efficiency after five

185   days. We calculated the percent of mNG positive cells and observed that a PE plasmid

186   concentration of 5 ng and MLH1dn plasmid concentration of 15.2 ng produced the highest

187   fraction of mNG positive cells (Fig. 5b). With the same MLH1dn condition, increasing the

188   amount of loaded PE plasmid decreased the fraction of mNG-positive cells (Fig. 5b). We

189   inspected our images for proper localization of mNG signal in cells and observed that it was

190   predominantly in the cytoplasm (Fig. 5c), consistent with previous reports of mNG2$_{11}$ tagging of

191   beta-actin(14). We noticed fewer total cells in conditions with high concentrations of MLH1dn.

192   To quantify this effect, we counted the number of Cellpose segments for each condition and

193   normalized it to the no plasmid control. We observed that MLH1dn inhibited cell proliferation in a

194   dose-dependent manner, with the lowest relative cell numbers observed for the highest

195   concentration of MLH1dn (Fig. 5d). For endogenous gene tagging, the total number of mNG-

196   positive cells is a practical metric to assess for cell editing throughput as it integrates editing

197   efficiency and cell proliferation. We counted the total number of mNG positive cells for each

198   condition and found that a lower MLH1dn concentration of 4.5 ng yielded the most mNG positive

199   cells, with PE concentration having little effect on the total number of mNG positive cells (Fig.

200   5e).

201

202   **Discussion**

203   Microscopy-based cytometry, as we demonstrated here, has multiple benefits. First, the

204   use of imaging approaches provides richer information per cell than flow cytometers, which can

205   be further analyzed after the fact, which is not the case with flow cytometry. Second, imaging

206   approaches allow for minimal handling of cells, only requiring a brief exposure to trypsin as

207   sample prep, which could be easily automated. Third, this approach only requires a simple

208   microscope, making it accessible to many researchers.

209   Microscopy-based cytometry relies heavily on the ability to segment cells from the

210   background and other cells accurately. While Cellpose may work well for many cell types and

211   conditions, particularly for fluorescent images, in our hands, it failed to reliably segment HEK

212   cells adhered to the surface using brightfield images only (Fig. 1a,d). This poor performance is

213   most likely due to the data type used to train the segmentation model and low contrast from

214   brightfield images of adherent cells. By inducing slight detachment of the cell from the surface

215   with trypsin, the contrast at the edges increased, allowing Cellpose to demark the boundaries of

216   cells more accurately (Fig. 1b,d). When cells were completely detached from the surface,

217   Cellpose can over-segment the images, providing more segments than real cells (Fig. 1c-d).

218   Cellpose performing better on only partially detached cells is consistent with the surface

219   adhered cell Cellpose was trained on as these partially adhered cells more resemble the

220   training data. The mechanical disruption and centrifugation of cells led to some cells being

221   stacked on each other, which could lead to misidentification due to out-of-focus fluorescence

222   contributing to adjacent cells (Fig. 3a 125K).

223   From a practical perspective, the preparation of trypsinized-only cells is more streamlined than

224   trypsinized and resuspended (also the typical workflow for flow cytometry), only requiring the

225   addition of trypsin and a quencher before imaging, decreasing the time and complexity of

226   experiments. A trypsinization-only method will maintain a cell's relative location to other cells.

227   This could allow colonies of cells to be assessed, facilitating a more complex analysis of many

228   cells from a clonal subpopulation, which could be informative to understanding heterogeneity in

229   biological processes. Additionally, because cells don't completely detach and ball up with a

230   trypsinization-only approach, if a nuclear stain is included, the cytoplasmic signal can easily be

231   differentiated from the nuclear signal, which is not possible when cells are trypsinized and

232   resuspended.

233

234

235   Cellpose parameter optimization demonstrated that proper selection of estimated

236   diameters is important for accurate discrimination of cell boundaries, whereas the flow threshold

237   had much less of an effect for the values we tested in our data (Fig. 2a-i). Cellpose does contain

238   functionality to estimate a diameter parameter instead of using a user-defined value but for our

239   data, Cellpose greatly underestimated the diameter and location of cells when using this

240   functionality. Cellpose estimated the average diameter of cells in our data to be 23.76 ± 3.25

241   pixels which is well below the 145 pixels we found to be optimal. We suggest users of this

242   method always be conscious of selecting an appropriate diameter parameters. Our observation

243   that Cellpose will over-segment images of trypsinized and resuspended cells if improper

244   parameters are chosen again suggests that the partially adhered cells more resemble the

245   training data as the maximum segments obtained from this group was the same as the number

246   of nuclei counted (Fig. 2b-c). For these reasons, we opted for the slight under-segmentation of

247   trypsinized cells instead of fully trypsinized and resuspended cells in our method. The false

248   negative and false positive ratios were calculated using the nuclei segments as the ground truth.

249   This is a limited method as the nucleus does not encompass all of the cytoplasm, particularly in

250   the adhered cells, but instead occupies 80-90% in our HEK cell images. For our purpose, these

251   metrics are sufficient to capture when multiple segments are called for the same cell and when

252   no segments are called for a cell.

253   We demonstrated that Cellpose faithfully segments all of the cells imaged up to 62.5k

254   cells/well (Fig. 3b), beyond which the cell density was too high, leading to cells stacking on top

255   of one another and underdetection. This demonstrates a cell density limitation to this approach,

256   as high density and cells that grow in more than a single layer might negatively affect the

257   accuracy of measurements. While we demonstrated that Cellpose works well for partially

258   adhered cells, it is also practical to use with non-adherent cells if the cells are allowed to settle

259   or be centrifuged to the surface. To overcome the limitation of Cellpose discussed, it is possible

260   to add annotations to a Cellpose model and improve its accuracy in a future version of

261   Cellpose(15). This could enhance segmentation in the case of non-adhered cells, but we found

262   the pre-trained models sufficient for our purposes.

263   We saw very few segments that contained a substantial amount of signal from both

264   mApple and mNG in our mixing experiments, which indicates a high level of single-cell purity

265   (Fig 4a-b). This is an important metric to assess an imaging approach especially when

266   comparing to flow cytometry. We suspected that cells in close proximity to each other would

267   contribute to the fluorescence measurement of neighboring cells but we do not see a substantial

268   influence of this. Our experiments comparing flow cytometry to our approach demonstrate that

269   imaging can yield higher SNR measurements than a conventional flow cytometry approach.

270   This is likely due to microscopes' optics, which are optimized better for light collection, and the

271   amount of exposure time in which signal can be collected. For flow cytometry, the exposure can

272   be on the order of less than a millisecond, while the exposure for microscopy is typically

273   hundreds of milliseconds. This longer exposure time increases the signal while the noise

274   remains relatively constant. While the laser power of the flow cytometer could always be

275   increased to compensate for shorter exposure, this may saturate the emitting ability of

276   fluorescence elements and could induce phototoxicity(16). Imaging-based approaches, as we

277   demonstrated, have the ability to measure lowly expressed proteins which can not be measured

278   on most flow cytometers. This has been shown to be particularly useful in assessing

279   endogenous protein tagging efficiency(17) and allows for a larger dynamic range in

280   measurements of fluorescent reporters.

281

282   A typical flow cytometry experiment takes approximately 2 hours to process a 96-well

283   plate with 50,000 cells per well, which includes 0.5 hours for the cell preparation and 1.5 hours

284   of flow cytometer time. The throughput of microscopy-based cytometry largely depends on the

285   size of the microscope field of view and the automation of the microscope. At a maximum cell

286   density of 62.5k cells per well of a 96-well plate (~200k cells/cm2), using 10X magnification on a

287   typical sCMOS camera (~ 1.3 x 1.3 mm2 field of view), one image will contain approximately

288   3,300 cells therefore requiring 16 images per well to obtain the same number of cells as a flow

289   cytometer. A standard microscope with a motorized stage can easily image a full 96-well plate in

290   1.5 hours, closely matching the throughput of a flow cytometer. Additionally, due to the high

291   level of coincident when high flow rates are used in flow cytometry, many events are discarded

292   due to doublet formation increasing the time and number of cells required to complete a flow

293   cytometry experiment.

294   We demonstrated the utility of our approach by screening a set of 88 prime editing conditions.

295   We found an optimal concentration for the amount of PE and MLH1dn to use when tagging

296   beta-actin with mNG$_{11}$. The conditions with the highest fraction of mNG-positive cells (Fig. 5b)

297   were ones with high levels of MLH1dn plasmids as well as some of the lowest cell proliferation

298   (Fig. 5d). This slow cell growth is not surprising as MLH1 is involved in meiotic crossing over

299   during cell proliferation(18) and MMR processes. While MMR inhibition with MLH1dn may

300   promote prime editing even at low concentrations of PE plasmid, its impact on cell viability

301   needs to be considered in prime editing optimization. We evaluated a more inclusive metric that

302   considered both editing efficiency and cell proliferation (Fig. 5e) as it is more relevant to

303   endogenous gene tagging methods with the ultimate goal of sorting cells into a pure population

304   of tagged cells. This is much more efficient with more tagged cells in total. Here, we screened

305   PE and MLH1dn plasmid concentrations but this method could easily be used to screen other

306   relevant factors in prime editing such as pegRNA sequence, pegRNA targeting site, and novel

307   PE architectures.

308   Beyond prime editing, this approach has utility in any assay measuring a fluorescent reporter

309   that one would typically be assayed on a flow cytometer. With a better SNR, simpler sample

310   preparation compared to flow cytometry, and richer data, even on simple microscopes, we

311   envision this approach replacing flow cytometry for diagnostic assays and in situations where

312   flow cytometers are not accessible. Additionally, this method is particularly useful where many

313   samples require assaying, and automated sampling for a flow cytometer is not available as

314   sample preparation is easily parallelizable.

315

316   **Methods**

317   **Cell culture**

318   HEK293 cells were cultured in DMEM + 10% FBS & 1% Penicillin-Streptomycin. Cells were

319   washed twice with 1x PBS, then detached by incubating 2-5 minutes at room temperature with 3

320   mL of 0.25% trypsin. Cells were seeded on glass bottom 96-well plates. For Cellpose validation

321   experiments, cells were stained with 10 µg/ml of Hoechst in PBS for 10 minutes. To induce

322   partial detachment of cells facilitating better segmentation, cells were first washed with PBS and

323   then treated with 0.25% trypsin for 3 minutes at 37°C. trypsin was quenched with DMEM + 10%

324   FBS and replaced with FluoroBrite DMEM + 10% FBS for imaging. For the trypsinized and

325   resuspended condition, cells were treated with 0.25% trypsin for 3 minutes at 37°C, quenched

326   with DMEM + 10% FBS, and mechanically agitated by repeatedly mixing via pipetting the

327   solution in the wells. Cells were then centrifuged in the wells for 3 minutes at 300 g to place

328   cells in the imaging plane. The media/trypsin solution was gently removed and replaced with

329   FluoroBrite DMEM + 10% FBS before imaging. For cell dilution experiments, HEK cells in

330   suspension were counted and diluted to appropriate concentrations such that 500000, 250000,

331   125000, 62500, 31250, or 15625 cells were placed in each well. Cells were then centrifuged in

332   the wells for 3 minutes at 300 g to place cells in the imaging plane. The media/trypsin solution

333   was gently removed and replaced with FluoroBrite DMEM + 10% FBS before imaging. For cell

334   mixing experiments, cells were mixed at appropriate ratio before being seeded and allowed to

335   appear and grow for two days.

336

337   **Image acquisition**

13

338 Samples were imaged with a widefield Nikon Ti-E microscope with a motorized stage, a

339 Hamamatsu ORCA Flash 4.0 camera, an LED light source (Excelitas X-Cite XLED1), and a 60X

340 CFI Plan Apo IR water immersion objective or a plan fluor 20X/0.45 air objective.

341 For SNR and CV comparisons, samples were also imaged with an EVOS Cell Imaging System

342 (Thermo Fisher: AMAFD1000), RFP (Thermo Fisher: AMEP4652) EVOS LED filter cube, and a

343 Plan Fluor 10X/0.3 air objective.

344

345 **Cellpose segmentation**

346 Brightfield images were segmented with Cellpose (4) using the general cytoplasmic model

347 described in the original publication. Segments on the edge of the images were removed as well

348 as segments above or below thresholds for cell size. These thresholds were set by first

349 discarding very small segments, in our case segments less than 2000 pixels for 60x images,

350 then fitting a normal distribution to the remaining segment sizes and setting the size thresholds

351 as two standard deviations away from the estimated mean in both directions. A similar approach

352 was applied to fluorescent nuclei images using the nuclei model in Cellpose and these

353 segments were used as a reference for calculating the Cellpose segments/ nuclei count and the

354 false positive and false negative rates. The segmentation of nuclei with Cellpose is much more

355 robust as the nuclei model in Cellpose was trained entirely on fluorescence images of nuclei

356 whereas the cytoplasmic model was trained on both fluorescence and brightfield images.

357 Moreover, nuclei of adjacent cells are often spatially separated in the fluorescence images.

358

359 **Calculating Cellpose accuracy and optimal parameters**

360 To calculate the number of Cellpose segments/nuclei the ratio of the number of Cellpose

361 segments obtained from brightfield images after size thresholding was compared to the number

362 of segments obtained from nuclei images after size thresholding for each image in each

363 condition. To calculate the false negative ratio for Cellpose in each image, the number of

364   brightfield segments that fell within each nuclei segment were calculated. If there were no

365   brightfield segments, this segment was considered to be false negative. Similarly, to calculate

366   the false positive ratio, the number of nuclei segments that fell within each brightfield segment

367   were calculated. If there were no nuclei segments, this segment was considered a false

368   positive.

369   To optimize the parameters for Cellpose, four representative images were considered for

370   the adhered, trypsinized, and trypsinized/resuspended conditions. Cellpose was applied to all

371   12 images for each combination of 25 diameters ranging from 25 to 505 pixels and 21 flow

372   threshold values from 0.05 to 2.05 for a total of 525 parameter combinations. The

373   segments/nuclei, false positive ratio, and false negative ratios were calculated for each

374   condition and set of parameters.

375

376   **Fluorescence cell line generation**

377   To generate the landing pad cell line, HEK293 cells were transfected with three plasmids using

378   jetOPTIMUS according to the manufacturer's instructions in a 6 well plate with cells grown to

379   80% confluency. The plasmids used are as followed: 500 ng of a plasmid expressing Sniper2L

380   Cas9 (19), 500 ng of a plasmid expressing a single guide RNA targeting the safe harbor locus in

381   the first intron of the adeno-associated virus integration site 1 (AAVS1) (targeting sequence

382   GTTAATGTGGCTCTGGTTCT) (Supplementary materials), and 3000 ng of a donor plasmid

383   containing in order: 800 bp section of homology to the endogenous AAVS1 site, a splice

384   acceptor, a T2A site, an attB site of the large serine recombinase Pa01 (5), a TagBFP-NLS, a

385   T2A site, a puromycin resistance gene, a polyadenylation signal, and another 800bp section of

386   homology to the endogenous AAVS1 site (Supplementary materials). After 5 days of editing,

387   edited cells were enriched by continual exposure to 1ug/ml puromycin.

388

389   To integrate the mApple sequence into the AAVS1 site of landing pad cells, two plasmids were

390   transfected in a similar fashion as above; 3000 ng of a plasmid containing an attP site of the

391   Pa01 protein with an mApple expression cassette, and 500 ng of a plasmid to express the large

392   serine recombinase Pa01 to facilitate mApple integration (Addgene plasmid #193460).

393   Integration was allowed to proceed for 5 days before mApple+, TagBFP- cells were enriched

394   using fluorescence-activated cell sorting (FACS). HSPA8 mNeonGreen HeLa cells were

395   generated as described previously (14).

396

397   **Cell fluorescence analysis**

398   To quantify fluorescence signals from individual cells, the brightfield segmentation boundaries

399   from Cellpose were converted to a mask and applied to all channels to exclude all data outside

400   of segments, then for each unique segment, the sum of all the pixel values in each channel

401   were calculated and normalized to the total number of pixels. Cells were considered positive for

402   fluorescence signals in each channel if the average pixel intensity was greater than the 99.9%

403   percentile value calculated for the parent cell line with no fluorescence protein. SNR was

404   calculated by dividing the difference in mean fluorescence of mApple-positive and non-

405   fluorescent cells by the standard deviation of non-fluorescent cells. CV was calculated by

406   dividing the standard deviation of the fluorescence of mApple-positive cells by the difference in

407   mean fluorescence of mApple-positive and non-fluorescent cells.

408

409   **Flow Cytometry**

410   Flow cytometry was performed on a BD FACSAria II. The mNG signal was measured with a 488

411   nm laser and a 530/30 bandpass filter. The mApple signal was measured using the 561 nm

412   laser and a 610/20 bandpass filter. For detector voltage optimization, the photomultiplier tube

413   voltage was set at varying values and data collected for each voltage setting. 10,000 events

414   were recorded for each sample. Files in the .fcs format were exported from the BD FACS Aria II

415   and were analyzed in Python.

416

Prime editing optimization

418   HEK293 cells stably expressing the $mNG2_{1-10}$ fragment were seeded into a 96-well plate as

419   described above. All cells were transfected with 15 ng of plasmids expressing sgRNA

420   (protospacer: GAAGCCGGCCTTGCACATGC) and 15 ng of a plasmid expressing pegRNA

421   targeting the N terminus of the beta-actin gene carrying an insert of $mNG2_{11}$ (Supplementary

422   materials) For the combination of conditions, we screened 11 MLH1dn plasmid (Addgene

423   plasmid #174824) concentrations from 0 ng/well to 77 ng/well and 8 PE plasmid

424   (Supplementary materials) concentrations from 5ng/well to 85ng/well. JetOPTIMUS was used

425   for all transfections according to the manufacturer's instructions. Cells were trypsinized to

426   partially detached cells as described above and images were taken on day 5. Editing efficiency

427   was calculated by taking the ratio of mNG-positive cells to total cells. Cell survival ratios were

428   obtained by counting Cellpose segments compared to no plasmid controls.

429

Statistical analysis

431   When error bars are present, data represents >= 3 independent experiments. To determine

432   significance in the differences between groups analysis of variance (ANOVA) was performed

433   and if significant Tukey HSD test was performed to assess significance between groups.

434   significance was determined if $p < 0.05$

435

References

437

438   1. McKinnon, K. M. Flow Cytometry: An Overview. Curr. Protoc. Immunol. 120, 5.1.1-5.1.11

439   (2018).

17

440   2. Vicar, T. et al. Cell segmentation methods for label-free contrast microscopy: review and

441   comprehensive comparison. *BMC Bioinformatics.* 20, 360 (2019)

442   3. Thomas, R. M., John, J. A review on cell detection and segmentation in microscopic

443   images. *International Conference on Circuit ,Power and Computing Technologies*

444   *(ICCPCT), Kollam, India*, pp. 1-5 (2017)

445   4. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M. Cellpose: a generalist algorithm for

446   cellular segmentation. *Nat. Methods* 18, 100–106 (2021).

447   5. Durrant, M. G. et al. Systematic discovery of recombinases for efficient integration of

448   large DNA sequences into the human genome. *Nat. Biotech.* 41, 488–499 (2023)

449   6. Feng, S. et al. Improved split fluorescent proteins for endogenous protein labeling. *Nat.*

450   *Commun.* 8, 370 (2017).

451   7. Zhou, S., Feng, S., Brown, D. & Huang, B. Improved yellow-green split fluorescent

452   proteins for protein labeling and signal amplification. *PLOS ONE* 15, e0242592 (2020).

453   8. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks

454   or donor DNA. *Nature.* 576, 149–157 (2019).

455   9. Mathis, N. et al. Machine learning prediction of prime editing efficiency across diverse

456   chromatin contexts. *Nat. Biotechnol.* 1–8 (2024) doi:10.1038/s41587-024-02268-2.

457   10. Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat.*

458   *Biotechnol.* 40, 402–410 (2022).

459   11. Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome

460   manipulation. *Nat. Rev. Genet.* 24, 161–177 (2023).

461   12. Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants

462   of editing outcomes. *Cell* 184, 5635-5652.e29 (2021).

463   13. Kamiyama, D. et al. Versatile protein tagging in cells with split fluorescent protein. *Nat.*

464   *Commun.* 7, 11046 (2016).

18

465 14. Cho, N. H. et al. OpenCell: Endogenous tagging for the cartography of human cellular

466 organization. *Science* 375, eabi6983 (2022).

467 15. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods*

468 19, 1634–1641 (2022).

469 16. Icha, J., Weber, M., Waters, J.C., Norden, C. Phototoxicity in live fluorescence

470 microscopy, and how to avoid it. *BioEssays*. Aug 39, 8 (2017)

471 17. Leonetti, M. D., Sekine, S., Kamiyama, D., Weissman, J. S. & Huang, B. A scalable

472 strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl.*

473 *Acad. Sci.* 113, E3501–E3508 (2016).

474 18. Baker, S.M. et al. Involvement of mouse Mlh1 in DNA mismatch repair and meiotic

475 crossing over. *Nat Genet.* Jul;13(3):336-42 (1996)

476 19. Kim, Y. et al. Sniper2L is a high-fidelity Cas9 variant with high activity. *Nat. Chem. Biol.*

477 19, 972–980 (2023)

478

479

480

Acknowledgments

Author contributions

D.F. performed the experiments, analyzed the data, and contributed to the interpretation of the results. Y.K. performed flow cytometry experiments. K.Y. conceived and designed the proof-of-concept experiments. S.R. assisted in developing data analysis methods. D. F and B.H. conceived the project. B.H. and D.F. wrote the manuscript.

491   **Data availability**

492   A public git repository is available at https://github.com/BoHuangLab/dcyto containing

493   source code and a minimum analytical dataset sufficient to reproduce figures in the main

494   text.

495

496   **Competing Interests**

497   The authors declare no conflict of interest.

498

499   **Figure legends**

500   **Fig. 1 Trypsin treatment improves Cellpose segmentation of brightfield images for**

501   **adherent cells. a**. A representative brightfield image of adherent HEK cells with Cellpose

502   segment outlines highlighted, overlaid with a fluorescence image of the nuclear stain in blue

503   color. **b**. Same as **a** for trypsinized cells. **c**. Same as **b** for trypsinized and resuspended

504   cells. **d**. Cellpose segments per number of counted nuclei. **e**. False negative ratio of

505   Cellpose. **f**. False positive ratio of Cellpose. Error bars represent standard deviation across

506   4 independent experiments. * denotes $p < 0.05$ when comparing groups.

507

508   **Fig. 2 Cellpose parameter optimization**. **a-c**. Heat map of Cellpose segments per number

509   of counted nuclei for all combinations of estimated cell diameter and flow threshold

510   assessed for adhered (left), trypsinized (middle), and trypsinized + resuspended (right)

511   condition. **d-f**. Same as **a-c** for false negative ratio. **g-i**. Same as **a-c** for false positive ratio.

512

513   **Fig. 3 Cell density optimization for Cellpose segmentation**. **a**. Representative images

514   and Cellpose segment outlines for 15.6k, 31.3k, 62.5k, and 125k cells/well of a 96 well plate.

515   **b**. Cellpose segments obtained from 49 images for varying number of loaded cells. Dotted

20

516   line is the theoretical trend expected if Cellpose segments and number of loaded cells were

517   perfectly correlated.

518

519   **Fig. 5 Prime editing optimization. a**. Schematic of prime editing. PE and pegRNA

520   complexes to form an active editor, the editor is targeted to specific location in genome via

521   the pegRNA, the editor nicks the gDNA and pegRNA hybridizes to nicked strand, the

522   reverse transcriptase on the editor extends the gDNA which includes the insert, the

523   extended strand hybridizes to the opposite gDNA strand displacing the endogenous gDNA,

524   the endogenous gDNA flap is excised, and the mismatched edit and endogenous gDNA is

525   resolved. **b**. Heatmap of %mNG positive cells for all PE and MLH1dn conditions tested. **c**.

526   Representative image of mNG tagged beta-actin **d**. Heatmap of relative cell proliferation for

527   all PE and MLH1dn conditions tested. **e**. Heatmap of total mNG positive cells for all PE and

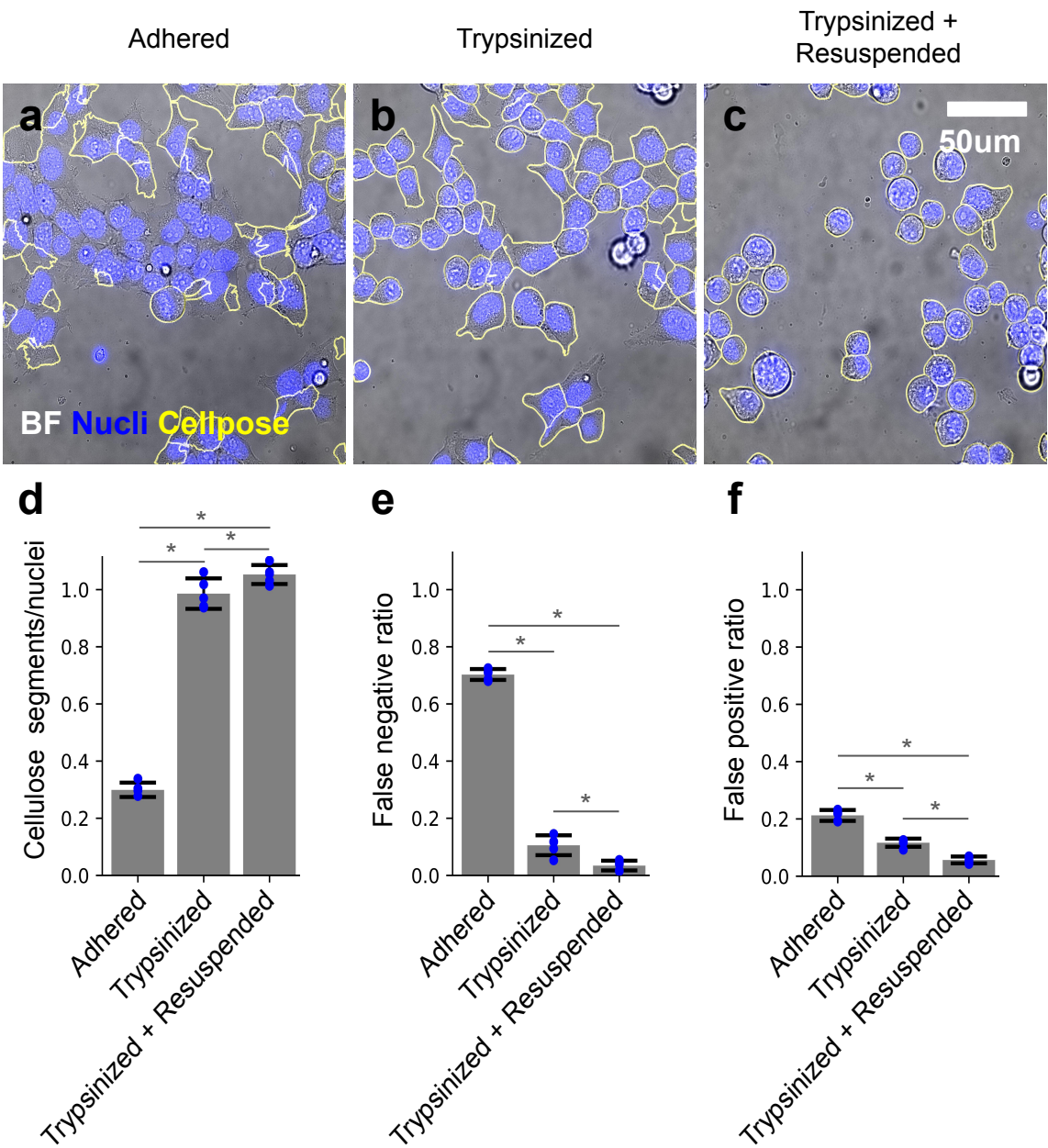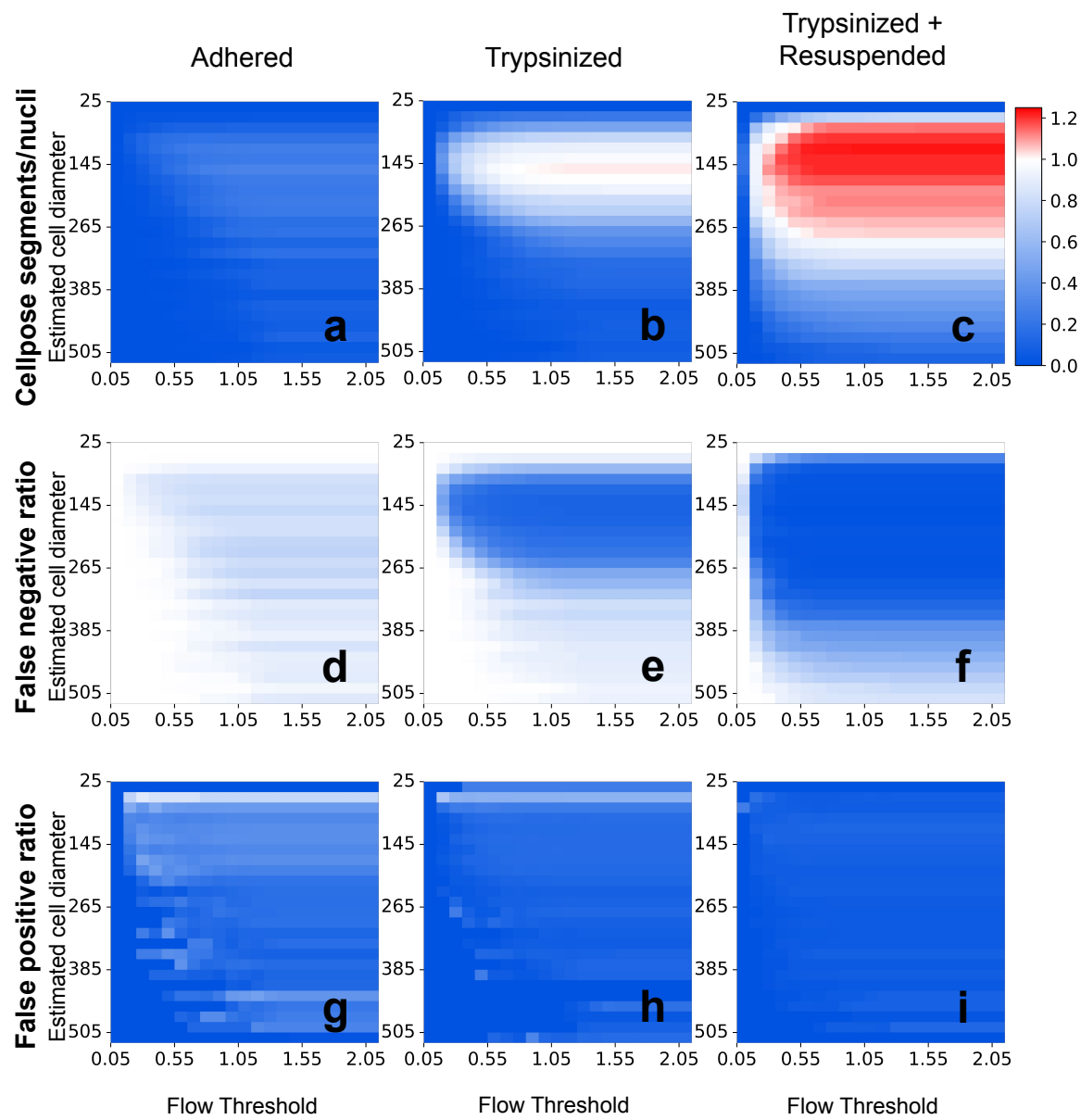528   MLH1dn conditions tested.

Figure 1

Figure 2
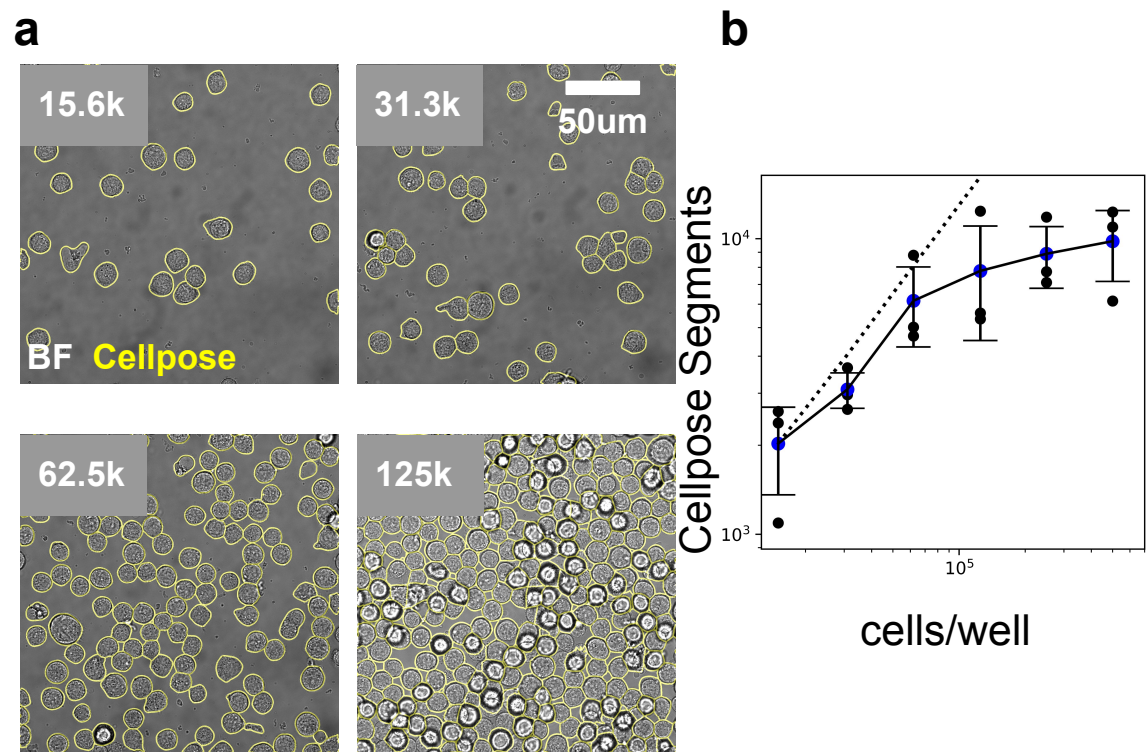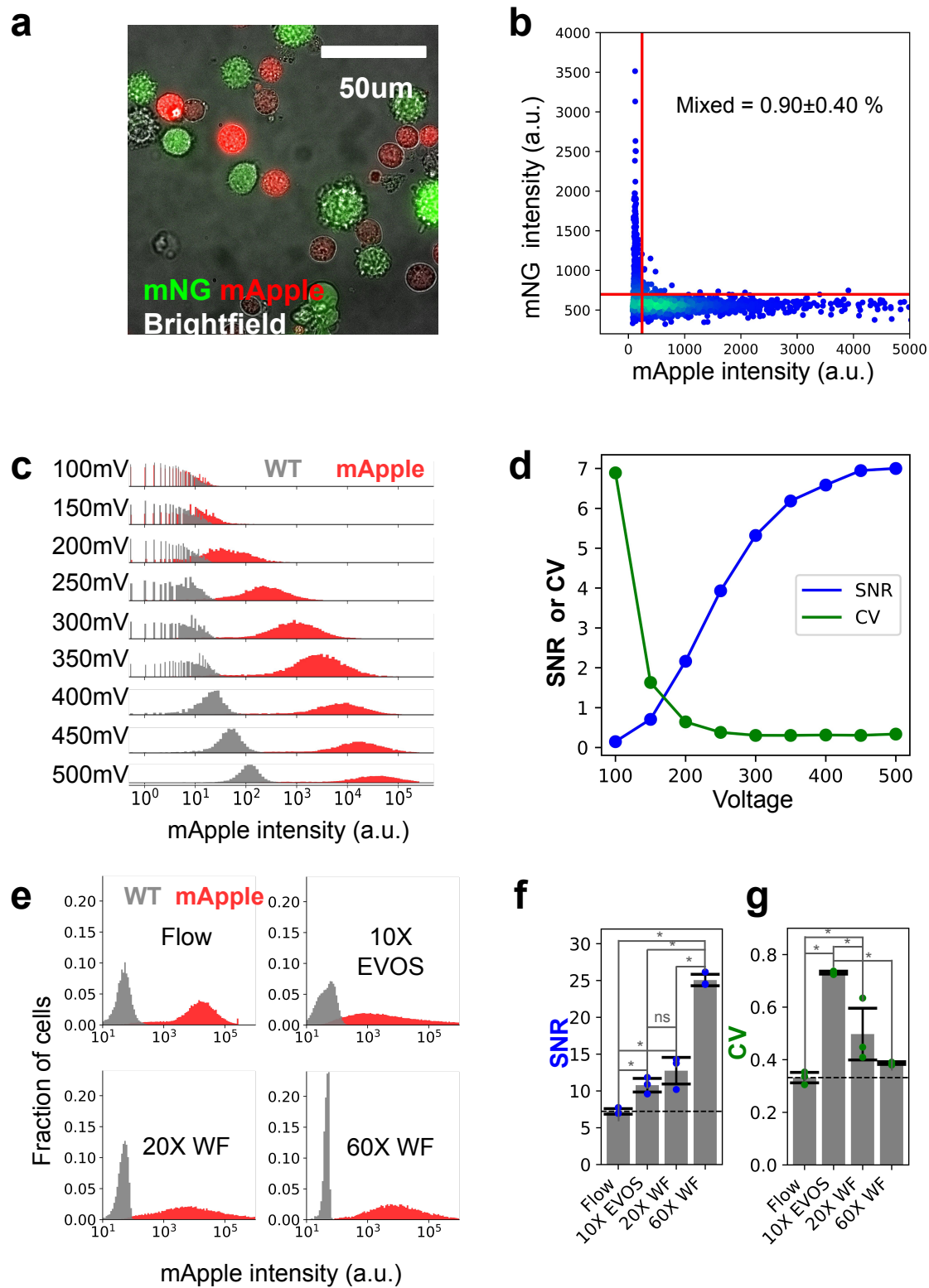
Figure 3

Figure 4

Figure 5