MORPHEUS: AN AUTOMATED TOOL FOR UNBIASED AND REPRODUCIBLE CELL MORPHOMETRY

(Article begins on next page)

MORPHEUS: AN AUTOMATED TOOL FOR UNBIASED AND REPRODUCIBLE CELL MORPHOMETRY

Federico Alessandro Ruffinatti[1]*, Tullio Genova[2,3]*, Federico Mussano[2] & Luca Munaron[3§]

[1] Department of Pharmaceutical Sciences, University of Eastern Piedmont, Novara, Italy
[2] CIR Dental School, Department of Surgical Sciences, University of Turin, Turin, Italy
[3] Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy

*These authors equally contributed; [§]Corresponding author

Keywords: cell morphometry; reproducibility; orientation analysis; Fiji; ImageJ.

[§] Correspondence to:
Luca Munaron, Ph.D.
Dept. Life Sciences & Systems Biology
University of Torino
Via Accademia Albertina 13
10123 Torino
ITALY
luca.munaron@unito.it

ABSTRACT

Here we present MORPHEUS, a new Fiji/ImageJ2 plugin for the automated evaluation of cell morphometry from images acquired by fluorescence microscopy. MORPHEUS works with sampling distributions to learn—in an unsupervised manner and by a nonparametric approach—how to recognize the cells suitable for subsequent analysis. Afterwards, the algorithm performs the evaluation of the most relevant cell-shape descriptors over the full set of detected cells. Optionally, also the extraction of nucleus features and a double-scale analysis of orientation can be performed. The whole algorithm is implemented as a one-click procedure, thus minimizing the user's intervention. By reducing biases and errors of human origin, MORPHEUS is intended to be a useful tool to enhance reproducibility in bioimage analysis.

INTRODUCTION

Cell morphometry can be defined as the quantitative description of cells' shape, which may not be separated from the inference of important information about their functional state. This shape–function paradigm aims to correlate cell morphology to biological processes such as proliferation, differentiation, migration, growth dynamics, adhesion to particular substrates or transition from physiology to pathology (Lepekhin et al., 2001; Ruffinatti et al., 2013; Yang et al., 2015; Lyons et al., 2016). Although many shape descriptors have been proposed over the years, no general rules about their usage currently exist since their suitability depends on the particular experimental aim (Chen et al., 2012; Lobo et al., 2015). More generally, cell morphometry lacks standardized protocols for the comparison of studies from independent laboratories. Specifically, when cells are manually selected, a heavy operator-dependent bias is introduced, impairing procedure reproducibility (Chen et al., 2016; Van Valen et al., 2016). In addition, since many samples are needed for reaching a good statistical power, human errors unavoidably arise because of the repetitiveness of such a time-consuming task.

47   In order to solve most of the aforementioned problems, we developed a new plug-in for Fiji

48   (ImageJ2) (https://imagej.net/ImageJ2/; https://fiji.sc/) called MORPHEUS (**M**ultipametric

49   **M**orphometric Analysis of **EU**caryotic cell**S)**. It combines within a single structured pipeline the

50   evaluation of the most relevant cell-shape descriptors, the extraction of nucleus features and an

51   advanced analysis of orientation. The procedure we propose is highly automated and adaptive since

52   the full algorithm is indeed implemented in a one-click and user-friendly fashion. MORPHEUS is

53   freely available for download at https://sourceforge.net/projects/morpheus-for-fiji/, together with

54   the same sets of test images used in Supplementary Materials for experimental validation (see

55   below).

56

57   **METHODS**

58   **Sample Preparation**

59   The following recommendations are made in order to generate MORPHEUS-compliant

60   images.

61   1. Cells stained for some cytoplasmic element (e.g. phalloidin for cytoskeleton) need to

62   be acquired by fluorescence microscopy. Cells stained for nuclei (e.g. through 4',6-

63   diamidino-2-phenylindole aka DAPI) are optional but not sufficient alone.

64   2. Images should be of good quality. If the Signal-to-Noise Ratio (SNR) is too low (i.e.

65   less than 10), the segmentation process will possibly fail (see Section 1.2 in

66   Supplementary Materials for an operational definition of SNR).

67   3. Both 8 and 16-bit gray scale images can be used, but MORPHEUS segmentation

68   procedure always assumes light objects (fluorescence signal) on dark background.

69   Different color lookup tables (LUTs) are not suitable for MORPHEUS.

70   4. Moderate and homogeneous plating densities are required. In particular, they ought

71   to be low enough to provide a majority of isolated cells compared to cell clusters. It

72   is warmly suggested to ensure even distribution of adherent cells by repetitive

73   pipetting.

74   5. Since all cell clusters are automatically excluded by the pipeline, images are not to

75   contain many pronounced cell elongations leading to cell-cell contacts (i.e. neuron

76   networks analysis or quantitative assays of dendritic branching and arborization).

77   Other platforms are specifically tailored for this aim (Ventimiglia et al., 1995; Wu et

78   al., 2004).

79   6. The number of cells per field should be large enough to produce low-skewness

80   sampling distributions for the statistics of interest (see below), otherwise MORPHEUS

81   will undergo an inaccurate learning step. Hence, the use of high magnification

82   objectives, leading to less than 10-15 cells per field, is discouraged.

83   The above limits are not expected to prevent the use of MORPHEUS in a great variety of

84   morphometric analysis contexts.

Algorithm Flow

Preliminary Steps

87   MORPHEUS can be easily integrated into Fiji *Plugins* menu by copying *Morpheus_.ijm* file into

88   the appropriate directory. When launched, a dialog window allows the user to set few basic

89   parameters (Figure 1a), among which an anti-*spot noise* threshold ε and a tolerance value T (the

90   larger T, the more inclusive the cell selection will be). After pressing OK button, no further

91   intervention is required by the user. As a preliminary task, MORPHEUS scans the input folder

92   searching for the images to be analyzed and then it performs a segmentation in five steps: contrast

93   enhancement, background subtraction, smoothing, thresholding and binarization (Figure 1b).

94   Please refer to Supplementary Materials for more details about the segmentation algorithm and the

95   other options of the starting dialog window.

96   *Unsupervised Learning Step*

97   As Fiji *Analyze Particles* function selects the objects to be measured primarily on the basis of

98   their area (A) and their circularity (C), a suitable range for these two parameters can be defined to

99   discriminate the objects of interest (i.e. the isolated cells) from the other entities detected by the

100   segmentation (i.e. small debris and big cell-clusters). MORPHEUS fulfils this task by analyzing the

101   distributions of the two random variables A and C over the entire population and learning from the

102   dataset the interval within which most of the isolated cells are likely to fall.

103   Even under controlled plating conditions, the overall distribution of A is not easily

104   predictable from a theoretical point of view. This is due both to the intrinsic biological variability

105   among cells and, as already mentioned, to the presence of different subpopulations of *objects*

106   within the same cell culture: spot noise, cellular and extracellular-matrix debris (corresponding to

107   the high frequencies near the origin of the histogram of areas), isolated cells (the central body of

108   the distribution) and cell clusters of two or more cells (the heavy right tail of the distribution; Figure

109   1d).

110   First, spot noise is removed by imposing a threshold ε at the beginning of the algorithm (a

111   justification for the default value ε=200 pixel^2 is provided in Supplementary Materials), then *Analyze*

112   *Particles* function evaluates A and C for all the segmented objects that are larger than ε (Figure 1c).

113   To cope with presence of sparse cell clusters, central tendency for A should be measured by an

114   estimator robust to outliers—the median being the simplest one—that can be thought to be more

115   representative of the single cell then the cell clusters (Figure 1d). Clearly, this is true only under the

116   initial hypothesis of low-density plating (point 4 of *Sample Preparation* section), meaning that the

117   majority of cells are indeed isolated cells. After the computation of the object median area $M_{A,i}$ for

118   each sample image ($i = 1, 2, ..., m$), the sampling distribution of the sample medians can be drawn

119   (Figure 1e). This distribution, albeit not normal, is generally far less skewed than that of the starting

population, provided the sample size is large enough (i.e. at least 10—25 objects per image,
depending on the dispersion of the overall population, as detailed in *Statistical Validation* section
of the Supplementary Materials). By explicitly drawing such a sampling distribution from the
experimental data, both the mean and the dispersion of the random variable M_A can be evaluated
in a direct way, thus providing a reliable estimate of the population median area together with its
accuracy or standard error (SE). This procedure leads to the definition of a 'characteristic' or 'typical'
single cell area ($\overline{M}_A$) and its uncertainty ($\text{SE}_{\overline{M}}$). Starting from these assumptions,

$$A_{high} = 2 \cdot (\overline{M}_A + T \cdot \text{SE}_{\overline{M}})$$

is a conservative estimate for the predicted typical size of a two-cell cluster. In other words, if an
object has $A > A_{high}$, it can be legitimately suspected to be a complex of two (or more) cells.
According to a log-symmetric criterion, the lower bound can be defined as

$$A_{low} = \frac{1}{2} \cdot (\overline{M}_A - T \cdot \text{SE}_{\overline{M}})$$

and all the objects below this limit will be discarded as cell debris (or residual spot noise). Depending
on the tolerance value T selected from the main window of MORPHEUS, the width of $\left[A_{low}, A_{high}\right]$
range can be changed, allowing for a more permissive ($T \rightarrow 6$) or more strict ($T \rightarrow 1$) approach to
cell selection.

As for the area, also the circularity generally exhibits a complex distribution, possibly made
of several subpopulations, since cell clusters can originate very low C values, while debris and spot
noise typically have C values very close to 1. Even in this case, MORPHEUS' approach is based on
sampling distributions, albeit with some important difference compared to area determination. The
multiplicative factor 2, used to predict the expected size of a two-cell cluster, is clearly meaningless
when dealing with circularity and there are no other obvious multiplicative factors able to predict
the value of C for those objects differing from an isolated cell. For this reason, a more conservative
approach is adopted, the only aim being that of trimming the outliers. By evaluating circularity

144   extreme values for each of the m sample images, MORPHEUS can draw the two sampling

145   distributions of sample minimum ($\min_C$) and sample maximum ($\max_C$). Again, these two

146   distributions can be conveniently described by their mean ($\overline{\min}_C$, $\overline{\max}_C$) and related SEs. Finally, by

147   tuning the T parameter, it can be set the range of circularity values to which single cells are actually

148   supposed to belong, namely:

149
$$C_{low} = \overline{\min}_C + (\tau - T) \cdot \mathrm{SE}_{\overline{\min}} ,$$

150
$$C_{high} = \overline{\max}_C - (\tau - T) \cdot \mathrm{SE}_{\overline{\max}} ,$$

151   Where the constant $\tau = 6$ corresponds to the highest confidence level the user can choose for

152   parameter estimation in the current implementation of MORPHEUS (see details in Section 1.1 of the

153   Supplementary Materials).

154   Once esteemed, $[A_{low}, A_{high}]$ and $[C_{low}, C_{high}]$ intervals are fed as constraints to *Analyze*

155   *Particles* for a second run of the function over the whole dataset (Figure 1f). In other words, they

156   are used to define a sort of bivariate confidence interval for the area and the circularity of the *true*

157   isolated cell, but with no reference to any underlying distribution parameter. MORPHEUS represents

158   indeed a completely nonparametric approach to cell morphometry, since no normality assumption

159   is made for any of the distributions considered. Importantly, the anti-spot threshold value ε and the

160   tolerance discrete level T are the only two parameters the user has to choose, thus maximally

161   reducing the arbitrariness of the procedure and all the ensuing possible biases. In particular, the

162   effects produced by different choices of the tolerance level have been examined and reported in

163   Table SM3 of Supplementary Materials (*Experimental Validation – Tolerance Effect* section).

164   *Morphometry and Orientation Analysis*

165   Using the (A, C)-bivariate confidence interval as a filter, MORPHEUS can recognize isolated

166   cells in each sample image (Figure 1g). Here the algorithm forks (Figure 1h): the first branch—always

167   performed—is dedicated to cell morphometry, while the second one—optional—is devoted to

168   orientation analysis. For the morphometric task, the idea is to use a wide descriptor spectrum to

169   capture as many features as possible, preferring to reduce the dimensionality of the problem during

170   the follow-up analysis rather than to choose a particular descriptor 'a priori' (see Figure 1i-j and

171   *Experimental Validation – Morphometry* section in Supplementary Materials). In particular, 12

172   different shape descriptors are evaluated for each detected cell, encompassing all those indexes

173   natively provided by Fiji/ImageJ through the *Set Measurements* function (i.e. area, perimeter, best

174   fitting ellipse (BFE) major axis, BFE minor axis, BFE aspect ratio, BFE angle, circularity, roundness,

175   solidity, Feret's diameter, Feret's angle, and minimum caliper diameter, as detailed in Section 1.3 of

176   the Supplementary Materials).

177   For the second task, MORPHEUS leans on OrientationJ, a well-established Fiji/ImageJ plugin

178   for orientation and isotropy characterization (Rezakhaniha et al., 2012; Püspöki et al., 2016).

179   MORPHEUS checks for OrientationJ (version ≥ 2.0.2) and, if present, *OrientationJ Distribution*

180   function is called for an automated directional analysis of the whole dataset over two distinct scales:

181   cytoskeleton and whole-cell (Figure 1k). The analysis is performed locally by means of a sliding

182   Gaussian window with an arbitrary radius σ that is meant to be as close as possible to the structure

183   of interest, namely $\sigma = 1$ pixel for fine structure (cytoskeleton) analysis and

184    $$\sigma = \sqrt{\frac{\overline{M}_A}{\pi}},$$

185   for the whole-cell level, taking advantage of the 'typical cell area' as estimated in the initial learning

186   step. For both the levels of analysis, MORPHEUS returns different outputs of both qualitative and

187   quantitative nature. In particular, two colormaps are saved carrying the directional information

188   according to a Hue-Saturation-Brightness (HSB) color model, where hue encodes for the local

189   dominant orientation, saturation encodes for the coherency and brightness is based on the gray

190   levels of the input image (Figure 1l). In addition, a set of coherency-weighted orientation histograms

191   are collected and then assembled into a single heatmap (Figure 1m).

192   *MORPHEUS Log*

193   With the aim of being as compliant as possible with the reproducibility standards in

194   bioinformatics (see e.g. Sandve et al., 2013), MORPHEUS prints an onscreen log at runtime and saves

195   a copy of it in the output directory. Both the onscreen log and log file contain all the information

196   needed to reproduce that particular analysis at a later time (see Supplementary Materials for

197   details).

198

199   **RESULTS**

200   To have a thorough experimental validation of MORPHEUS, the interested reader can refer

201   to the *Experimental Validation* section of the Supplementary Materials, in which a detailed analysis

202   of a prototypical MORPHEUS output is presented applying the algorithm to four different original

203   datasets of fluorescent cell samples.

204

205   **DISCUSSION**

206   Replication of experimental results at all scales of biological sciences is a pillar of scientific

207   method. However, the *crisis of reproducibility* or *replication crisis* has been largely debated on both

208   generalist and more specialized journals in the last years (Aleksandra et al., 2018; An, 2018; Baker,

209   2016; Begley & Ellis, 2012; Coiera et al., 2018; Fanelli, 2018; França & Monserrat, 2018; Samsa &

210   Samsa, 2019; Sandve et al., 2013). Different and variable factors contribute for this deep limitation

211   in modern wet biology as well as bioinformatics fields: among them, low statistical power, poor

212   analytic approaches and a lack of standardization in experimental protocols are unfortunately quite

213   usual in preclinical and clinical research papers.

According to such an increasing requirement for quality assessment and robustness, MORPHEUS is a promising proposal for the standardization of cell morphometric analysis. Importantly, the automation descending from the adaptive nature of the algorithm eliminates the user-bias factor and translates into a reproducibility-oriented approach to cell morphometry. This is intended to avoid all those human errors typically arising from a repetitive and time-consuming task, as well as the systematic bias and arbitrariness introduced by the experimenter whenever a threshold needs to be chosen or several objects are to be recognized by sight. A general outcome of the algorithm presented in this paper lies in the intriguing inversion of the usual time spent by the operator to obtain the final results. Indeed, in the canonical 'manual' approach most of the time is dedicated to the morphometric analysis of any single frame, leading to a significant limitation in the number of processed images. Otherwise, since MORPHEUS dramatically speeds up the analysis (virtually instant output), the operator will hopefully devote most of the experimental effort in image acquisition, thus increasing the quality and the statistical power of the work.

REFERENCES

Aleksandra, T., George, L., Velocity, H. H., & Jonas, R. (2018). Statistical mistakes and how to avoid them – lessons learned from the reproducibility crisis. *Osteoarthritis and Cartilage*, 1–3.

An, G. (2018). The Crisis of Reproducibility, the Denominator Problem and the Scientific Role of Multi-scale Modeling. *Bulletin of Mathematical Biology*, *80*(12), 3071–3080.

Baker, M. (2016). Is there a reproducibility crisis? *Nature*, *533*(7604), 452–454.

Begley, C. G., & Ellis, L. (2012). preclinical cancer research. *Nature*, *483*, 531–533.

Chen,S. et al. (2012) Recent advances in morphological cell image analysis. Comput. Math. Methods. Med., 2012, 101536.

Chen, C. L., Mahjoubfar, A., Tai, L. C., Blaby, I. K., Huang, A., Niazi, K. R., & Jalali, B. (2016). Deep Learning in Label-free Cell Classification. *Scientific Reports*, *6*(March), 1–16.

Coiera, E., Ammenwerth, E., Georgiou, A., & Magrabi, F. (2018). Does health informatics have a replication crisis? *Journal of the American Medical Informatics Association*, *25*(April), 963–968.

Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, *115*(11), 2628–2631.

França, T. F., & Monserrat, J. M. (2018). Reproducibility crisis in science or unrealistic expectations? *EMBO*

244   *Reports*, *19*(6), e46008.

245   Lepekhin, E. A., Walmod, P. S., Berezin, A., Berezin, V., & Bock, E. (2001). Evaluation of Cell Morphology. In R.
246   H. Gavin (Ed.), *Cytoskeleton Methods and Protocols* (pp. 85–100). Totowa, NJ: Humana Press.

247   Lobo, J., See, E. Y.-S., Biggs, M., & Pandit, A. (2016). An insight into morphometric descriptors of cell shape
248   that pertain to regenerative medicine. *Journal of Tissue Engineering and Regenerative Medicine*, *10*(7),
249   539–553.

250   Lyons, S. M., Alizadeh, E., Mannheimer, J., Schuamberg, K., Castle, J., Schroder, B., … Prasad, A. (2016).
251   Changes in cell shape are correlated with metastatic potential in murine and human osteosarcomas.
252   *Biology Open*, *5*(3), 289–299.

253   Püspöki, Z., Storath, M., Sage, D., & Unser, M. (2016). Transforms and Operators for Directional Bioimage
254   Analysis: A Survey. In Springer (Ed.), *Focus on Bio-image informatics* (pp. 69–93).

255   Rezakhaniha, R., Agianniotis, A., Schrauwen, J. T. C., Griffa, A., Sage, D., Bouten, C. V. C., … Stergiopulos, N.
256   (2012). Experimental investigation of collagen waviness and orientation in the arterial adventitia using
257   confocal laser scanning microscopy. *Biomechanics and Modeling in Mechanobiology*, *11*(3–4), 461–473.

258   Ruffinatti FA, Gilardino A, Lovisolo D, Ferraro M. Spatial wavelet analysis of calcium oscillations in developing
259   neurons. PLoS One. 2013;8(10):e75986.

260   Samsa, G., & Samsa, L. (2019). A Guide to Reproducibility in Preclinical Research. *Academic Medicine*, *94*(1),
261   47–52.

262   Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten Simple Rules for Reproducible Computational,
263   *9*(10), 1–4.

264   Van Valen, D. A., Kudo, T., Lane, K. M., Macklin, D. N., Quach, N. T., DeFelice, M. M., … Covert, M. W. (2016).
265   Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments.
266   *PLoS Computational Biology*, *12*(11), 1–24.

267   Ventimiglia, R., Jones, B. E., & Møller, A. (1995). A quantitative method for morphometric analysis in neuronal
268   cell culture: unbiased estimation of neuron area and number of branch points. *Journal of Neuroscience*
269   *Methods*, *57*(1), 63–66.

270   Wu, C.-C., Reilly, J. F., Young, W. G., Morrison, J. H., & Bloom, F. E. (2004). High-throughput Morphometric
271   Analysis of Individual Neurons. *Cerebral Cortex*, *14*(5), 543–554.

272   Yang W, Tian R, Xue T. Nuclear shape descriptors by automated morphometry may distinguish aggressive
273   variants of squamous cell carcinoma from relatively benign skin proliferative lesions: a pilot study.
274   Tumor Biol. 2015;36(8):6125–31.

275