

¹ SCEMATK: Single Cell Morphological Profiling in Whole Slide Images

³ **Hugh Warden**  ¹, **Jesko Wagner**  ¹, and **Ava Khamseh**  ^{1,2}¶

⁴ ¹ MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK
⁵ ² School of Mathematics, University of Edinburgh, Edinburgh, UK ¶ Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Open Journals](#) ↗

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

⁶ Summary

⁷ Digital pathology is an important field of study for fast and accurate diagnostic processes as well as for structural analysis of diseased tissue. These large high resolution images are then shared amongst pathologists and annotated, however recently it there has been a lot of work to apply machine learning to these images to perform common diagnostic and investigative work. Due to the large size of the images, this requires a lot of technical expertise and computational resources to open and analyse these images. SCEMATK (Single Cell Extraction and Morphological Analysis Tool-Kit) provides an efficient way to open and convert digital pathology images into a phenotypic data set of morphological features that can then have machine learning techniques applied to them that do not take images as input. Furthermore, SCEMATK focuses on the processing patches uniformly, allowing for the spatial analysis of the morphology of cells between patches to capture greater information about larger tissue structure. All of these processes can be run interactively in a Jupyter notebook, allowing for easy manipulation of images too large to fit into memory.

²⁰ Statement of Need

²¹ Due to the large size of digital pathology whole slide images (WSIs), substantial computational resources are required to load, process, and analyse these data. Existing open source tools typically do not focus primarily on the morphological profiling of these images. QuPath provides comprehensive tools for image viewing, annotation and cell detection [cite{QuPath}]. It can perform some morphological profiling however more complex analytical workflows require a steep learning curve. Other software focuses on preparing WSIs for machine learning using efficient tiling and patch extraction strategies [cite{TIAToolbox, HistomicsTK, SlideFlow, SlideTiler}]. Although some of these tools support morphological profiling, the independent processing of tiles can introduce inconsistencies when calculating spatial statistics.

³⁰ SCEMATK addresses this gap by providing a unified, easily extensible interface for image preprocessing, object segmentation, and morphological profiling, implemented using lazy-loading strategies. This design enables efficient analysis on standard personal computers while remaining scalable to high-performance computing environments. In contrast to existing morphological profiling tools, SCEMATK applies image transformations consistently across the entire slide prior to feature extraction, allowing computation of spatial statistics that characterise the local distribution of phenotypes within each cell's neighbourhood.

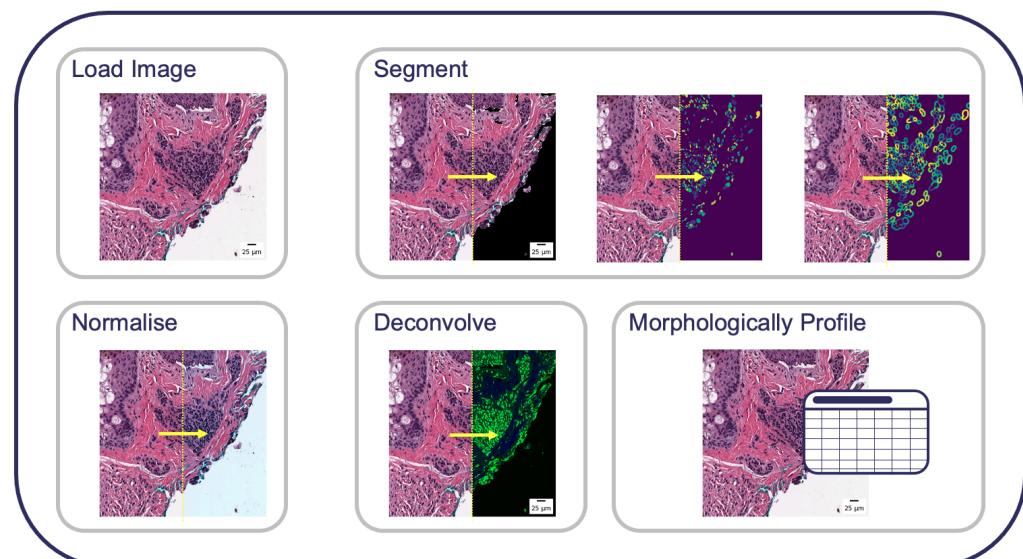


Figure 1: Overview of scematk

37 SCEMATK provides these basic tools for analysing images:

- 38 ▪ Image Viewing: Enables the user to view a thumbnail of an image or a specify a region
39 to view in full resolution.
- 40 ▪ Image Normalisation: Apply a colour transform to the image to fit that of a target image,
41 combatting batch effects that may arise from differences in staining protocols between
42 samples.
- 43 ▪ Stain Deconvolution: Convert the image from RGB to a new colour space showing
44 the intensities of different stains applied to the sample, indicating different biological
45 properties of the tissue.
- 46 ▪ Object Segmentation: Identify nuclei of cells within an image as well as an approximation
47 of the cell and cytoplasm.
- 48 ▪ Morphological Profiling: Convert the masks and deconvolved image into a table of
49 morphological properties of each cell.

50 SCEMATK provides an all in one easy to use interface for carrying out single cell morphological
51 profiling on digital pathology images with limited computational resources.

52 Acknowledgements

53 HW and JW are funded by an MRC Unit Award.