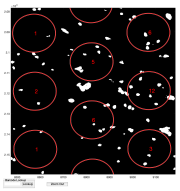
Preprint
http://research.libd.org/VisiumLIBD
Advance Access Publication Date: Day Month Year
Application Note



Visium Histology Image Processing

VistoSeg: a MATLAB pipeline to process, analyze and visualize high-resolution histology images for Visium spatial transcriptomics data

Madhavi Tippani [1,*], Heena R. Divecha [1], Joseph L. Catallini II [1,2], Lukas M. Weber [2], Abby Spangler [1], Andrew E. Jaffe [1,2,4], Stephanie C. Hicks [2], Keri Martinowich [1,3,4], Leonardo Collado-Torres [1], Stephanie C. Page [1], Kristen R. Maynard [1,3,*]

[1]Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD 21205, USA
[2]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
[3]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[4]Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Recent advances in spatially-resolved transcriptomics technologies such as 10x Genomics Visium platform have enabled the generation of transcriptome-wide spatial expression maps within intact tissue. However, steps for processing the high-resolution histology images, extracting relevant features from the images, and integrating them with the gene expression data remain unresolved.
Results: We describe *VistoSeg*, a MATLAB pipeline to process, analyze, and interactively visualize the high-resolution images from the 10x Genomics Visium platform. The output from *VistoSeg* can then be integrated with the spatial-molecular information in downstream analyses using any programming language, such as R or Python.
Availability: *VistoSeg* is available at https://github.com/LieberInstitute/VistoSeg (DOI: 10.5281/zenodo.5156783) with a tutorial at http://research.libd.org/VistoSeg
Contact: Kristen.Maynard@libd.org, Madhavi.Tippani@libd.org
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The 10x Genomics Visium workflow uses a spatial barcoding strategy to map RNA-sequencing reads to specific anatomical locations ('spots') on a histological image to create an integrated map of transcriptome-wide gene expression within an intact tissue section. Currently, 10x Genomics Loupe Browser (10x Genomics, 2021a) does not support analysis of histological images for integration with molecular gene expression data. However, these images represent a potentially rich, untapped resource for downstream analyses, such as to estimate the number of cells in each Visium spot and to identify cell type based on morphology (e.g. pyramid-shaped excitatory neurons). Analysis of histological images can identify spots containing a single cell or spots lacking cell bodies, which may represent specific cellular compartments (i.e. axon and dendrite-rich neuropil in brain tissue). Several software packages have been developed in the Python and R programming languages for cell segmentation and subsequent registration of microscopy images to anatomical reference atlases (Stringer *et al.*, 2021) (Fürth *et al.*, 2018). While some of these frameworks have been applied to spatially resolved transcriptomics data (Ortiz *et al.*, 2020), they were designed and optimized for cultured cells and mouse brain tissue sections. In contrast, other packages, such as Fiji (Schindelin *et al.*, 2012) and CellProfiler (McQuin *et al.*, 2018) lack

functions to split the slide images, and have large memory requirements. Furthermore, most High Performance Computing (HPC) systems do not fully support these memory-intensive interactive applications. To address these challenges, we developed *VistoSeg*, a MATLAB-based software package to facilitate segmentation, analysis, and visualization of Visium histological data derived from human brain for integration with gene expression data.

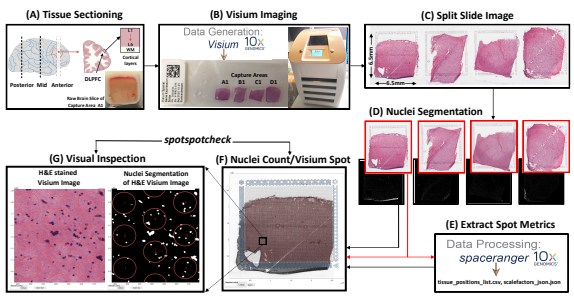


Fig. 1. An overview of VistoSeg. (A) Tissue sections from three levels of human dorsolateral prefrontal cortex (DLPFC; anterior, middle & posterior) spanning six neuronal layers plus white matter. (B) 10x Visium platform used to generate spatial-molecular data and histological images using Leica Aperio slide scanner. (C) A high-resolution slide image of four tissue sections to be split into individual images using `splitSlide()`. (D) Nuclei segmentations of individual images obtained from `VNS()` and `refineVNS()`. (E) Space Ranger output files that provide spot metrics. (F) The `countNuclei()` function counts the number of nuclei per spot. (E) The `spotspotcheck()` function enables interactive visualization of nuclei segmentations, provides options to zoom in and out, to search for spatial barcodes, and displays the nuclei count for all spots.

2 Results

VistoSeg is an automated MATLAB pipeline for processing high-resolution histological images for downstream spatial transcriptomics analysis. The 10x Visium platform is used to generate histological images from human brain tissue sections (**Fig. 1A**). As a first step, *VistoSeg* (i) processes these images by splitting them into the capture areas (currently, four) (**Fig. 1C**) corresponding to individual tissue sections using the `splitSlide()` function. The capture areas are saved as individual TIFF files at 70% resolution of the original raw image, as MATLAB cannot store data larger than $2^{32} - 1$ bytes in an image (TIFF, PNG, JPEG). Second, *VistoSeg* (ii) segments nuclei from TIFF files (**Fig. 1D**) using k-means color-based segmentation in the Image Processing Toolbox (The MathWorks, 2021) with the `VNS()` function (Visium Nuclei Segmentation). This function requires the user to input the total number of colors that are visually distinguishable from one another. Third, *VistoSeg* (iii) refines the binary images with the `refineVNS()` function for accurate detection of individual nuclei, i.e to split nuclei clusters. Fourth, *VistoSeg* (iv) uses the output TIFF files from (i) as input to the 10x Genomics Space Ranger software (10x Genomics, 2021b) for quantification of gene expression in a spatial context (**Fig. 1E**). This software generates a *CSV* and a *JSON* file that contains metrics to reconstruct the spot grid. (v) The `countNuclei()` function counts the number of nuclei per Visium spot in the corresponding image (**Fig. 1F**). Finally, (vi) the `spotspotcheck()` function (**Fig. 1G**) allows for interactive visual inspection of the segmentations from (ii). Using the output from Space Ranger, this function reconstructs and overlays the spot grid onto the histological image and binary segmentation to display the nuclei count. It allows the user to toggle between the nuclei segmentation and histology images, to search for spots with specific spatial barcode IDs, and to zoom in and out to a specific location on the image. Examples of code, output and results from publicly available datasets are available at http://research.libd.org/VistoSeg.

3 Methods

The high-resolution (0.253 microns per pixel) images are produced with a Leica Aperio slide scanner (**Fig. 1B**) equipped with a color camera and a 20X/0.75NA objective with a 2X optical magnification changer as described by 10x Genomics (10x Genomics, 2021c) for spatial transcriptomics.

4 Discussion

VistoSeg leverages the MATLAB Image Processing Toolbox (The MathWorks, 2021) to provide user-friendly functionality for processing, analysis, and interactive visualization of images generated from the 10x Genomics Visium platform. The output from *VistoSeg* can be integrated with spatial gene expression data for additional resolution and biological insights in downstream analyses. For example, using histological images of human DLPFC analyzed with VistoSeg, we identified "neuropil" spots lacking cell bodies, and demonstrated these spots are enriched for genes preferentially expressed in the transcriptome of synaptic terminals (Maynard *et al.*, 2021). *VistoSeg* currently has the following limitations: (i) large memory requirements for loading and saving images, and (ii) the `VNS()` function requires manual user inputs. Image processing solutions like *VistoSeg* enable the rich anatomical data in histological images to be integrated with corresponding spatial gene expression data and can serve as testing ground for ideas for future work. Future directions for *VistoSeg* include how to integrate machine learning strategies like VAMPIRE (Phillip *et al.*, 2021) to identify and classify cell types from the nuclei segmentations to be used for spot-level deconvolution of cell types.

Acknowledgements

Anthony Ramnauth (LIBD) for testing code functionality.

Funding

This work was supported by the Lieber Institute for Brain Development and National Institute of Health grant U01 MH122849 to KM.

References

10x Genomics (2021a). *10x Genomics Loupe Browser*.

10x Genomics (2021b). *10x Genomics Space Ranger Software*.

10x Genomics (2021c). *10x Genomics Visium Spatial Gene Expression Imaging Guidelines*.

Fürth, D. *et al.* (2018). An interactive framework for whole-brain maps at cellular resolution. *Nature neuroscience*, **21**(1), 139–149.

Maynard, K. R. *et al.* (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience*, **24**(3), 425–436.

McQuin, C. *et al.* (2018). Cellprofiler 3.0: Next-generation image processing for biology. *PLoS biology*, **16**(7), e2005970.

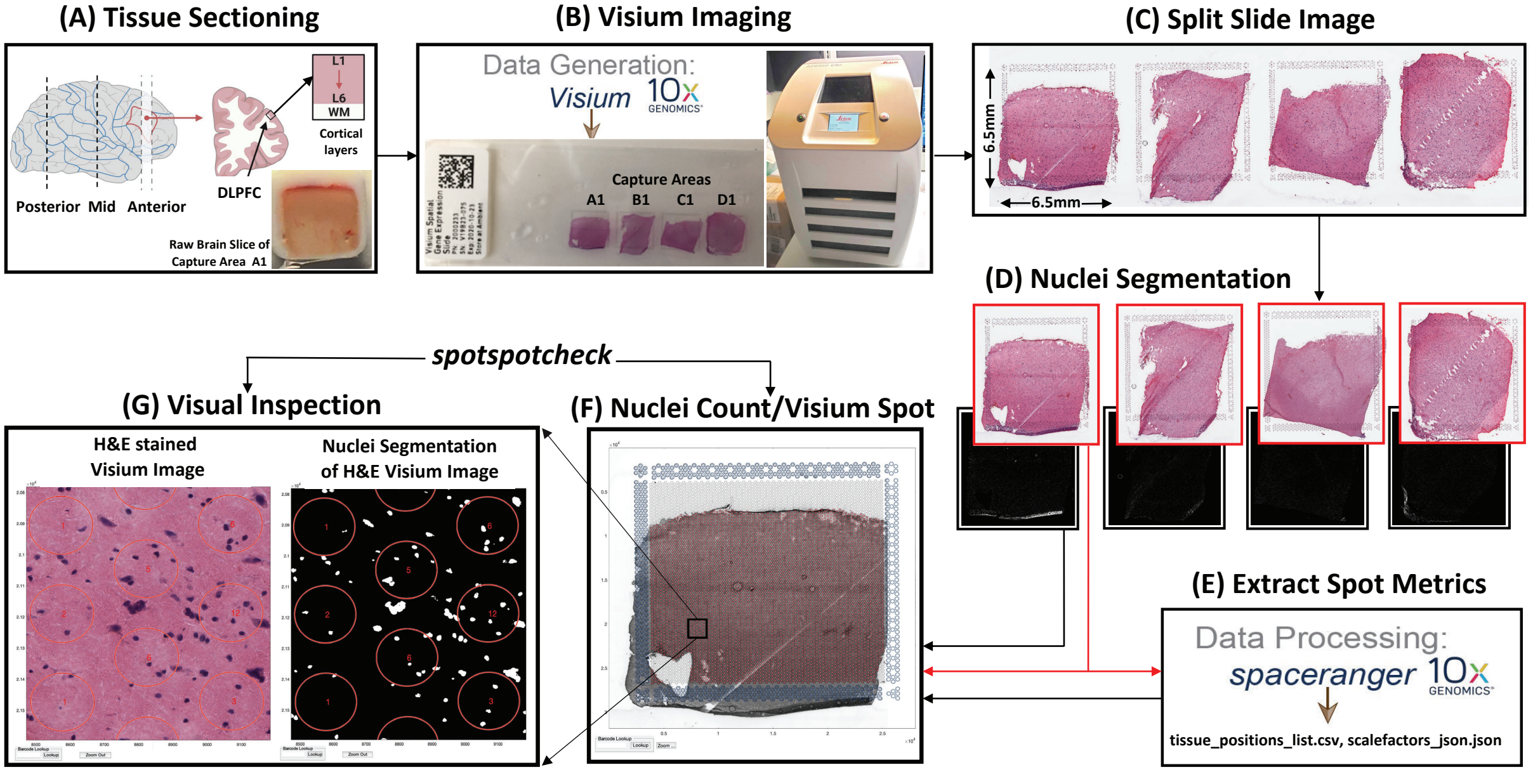
Ortiz, C. *et al.* (2020). Molecular atlas of the adult mouse brain. *Science Advances*, **6**(26), eabb3446.

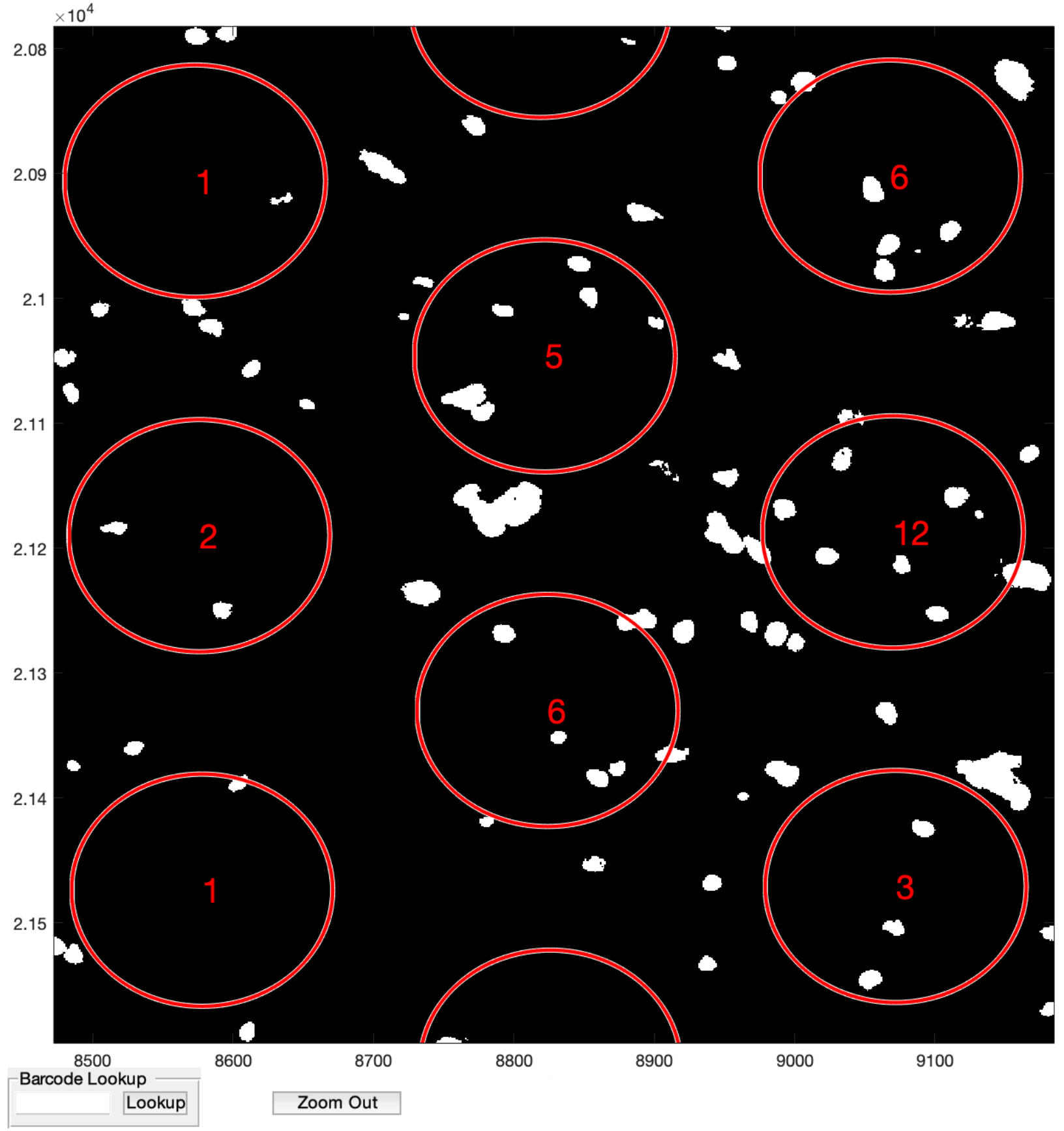
Phillip, J. M. *et al.* (2021). A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei. *Nature protocols*, **16**(2), 754–774.

Schindelin, J. *et al.* (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, **9**(7), 676–682.

Stringer, C. *et al.* (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, **18**(1), 100–106.

The MathWorks, I. (2021). *Image Processing Toolbox*. Natick, Massachusetts, United States.


(A) Tissue Sectioning
L1
L6
WM
Cortical layers
DLPFC
Posterior Mid Anterior
Raw Brain Slice of Capture Area A1
(B) Visium Imaging
Data Generation:
Visium
10x GENOMICS
Capture Areas
A1 B1 C1 D1
(C) Split Slide Image
6.5mm
6.5mm
(D) Nuclei Segmentation
spotspotcheck
(G) Visual Inspection
H&E stained Visium Image
Nuclei Segmentation of H&E Visium Image
(F) Nuclei Count/Visium Spot
(E) Extract Spot Metrics
Data Processing:
spaceranger
10x GENOMICS
tissue_positions_list.csv, scalefactors_json.json


×10^4
2.08
2.09
2.1
2.11
2.12
2.13
2.14
2.15
8500
8600
8700
8800
8900
9000
9100
1
6
5
2
12
6
1
3
Barcode Lookup
Lookup
Zoom Out