Journal of Visualized Experiments

Clonal analysis of the neonatal mouse heart using nearest neighbor modeling

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE61656R2
Full Title:	Clonal analysis of the neonatal mouse heart using nearest neighbor modeling
Section/Category:	JoVE Developmental Biology
Keywords:	regeneration, clonal analysis, lineage tracing, neonatal mouse, cryoinjury, heart
Corresponding Author:	Ravi Karra Duke University School of Medicine Durham, NC UNITED STATES
Corresponding Author's Institution:	Duke University School of Medicine
Corresponding Author E-Mail:	ravi.karra@duke.edu
Order of Authors:	Melanie Bakovic
	Devang Thakkar
	Paige DeBenedittis
	Diana C Chong
	Michael C Thomas
	Edwin S Iversen
	Ravi Karra
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Durham, NC, USA

Ravi Karra, MD MHS
Assistant Professor of Medicine
Division of Cardiology
Duke University Medical Center

July 20, 2020

Dear Dr. Upponi,

It is my pleasure to our revised article “Clonal analysis of the neonatal mouse heart using nearest neighbor modeling” to JOVE.

Included in this resubmission, you will find a revised manuscript, new supplemental coding files, and a point-by-point response to each of the editorial comments.

Thank you for the opportunity to submit our work to JOVE.

Sincerely,

Contact Information
Office: (919) 684-2449
Pager: (919) 970-5210
email: ravi.karra@duke.edu

Mailing Address
Box 3126 DUMC
Durham, NC 27710

Laboratory Address
Duke University Medical Center
213 Clinical Research Drive
CaRL Bldg., Rm 250
Durham, NC 27710

TITLE:

Clonal Analysis of the Neonatal Mouse Heart Using Nearest Neighbor Modeling

AUTHORS AND AFFILIATIONS:

Melanie Bakovic[1,*], Devang Thakkar[2,*], Paige DeBenedittis[1], Diana C. Chong[1], Michael C. Thomas[1], Edwin S. Iversen[3], Ravi Karra[1,4]

[1]Department of Medicine, Duke University, Durham, NC, USA
[2]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA
[3]Department of Statistical Science, Duke University, Durham, NC, USA
[4]Regeneration Next, Duke University, Durham, NC, USA

*These authors contributed equally.

Corresponding author:
Ravi Karra (ravi.karra@duke.edu)

Email addresses of co-authors:
Melanie Bakovic (melanie.bakovic@duke.edu)
Devang Thakkar (devang.thakkar@duke.edu)
Paige DeBenedittis (paige.debenedittis@duke.edu)
Diana C. Chong (dchong@gnf.org)
Michael Thomas (michael.c.thomas@duke.edu)
Edwin Iversen (iversen@duke.edu)

KEYWORDS:

regeneration, clonal analysis, lineage tracing, neonatal mouse, cryoinjury, heart

SUMMARY:

Presented here is a protocol for using multicolor lineage tracing and nearest-neighbor modeling to identify clonally derived cardiomyocytes during growth and regeneration in mice. This approach is objective, works across different labeling conditions, and can be adapted to incorporate a variety of image analysis pipelines.

ABSTRACT:

By replacing lost or dysfunctional myocardium, tissue regeneration is a promising approach to treat heart failure. However, the challenge of detecting bona fide heart regeneration limits the validation of potential regenerative factors. One method to detect new cardiomyocytes is the multicolor lineage tracing with clonal analysis. Clonal analysis experiments can be difficult to undertake, because labeling conditions that are too sparse lack sensitivity for rare events such as cardiomyocyte proliferation, and diffuse labeling limits the ability to resolve clones. Presented here is a protocol to undertake clonal analysis of the neonatal mouse heart by using statistical modeling of nearest neighbor distributions to resolve cardiomyocyte clones. This approach enables resolution of clones over a range of labeling conditions and provides a robust analytical

approach for quantifying cardiomyocyte proliferation and regeneration. This protocol can be adapted to other tissues and can be broadly used to study tissue regeneration.

INTRODUCTION:

A histologic hallmark of heart failure is the loss of cardiomyocytes (CMs), either following injury, senescence, or apoptosis[1]. Replenishing lost or dysfunctional myocardium through tissue regeneration represents a potential therapeutic strategy for curing patients with heart failure. Over the past several decades, seminal advances in developmental and regenerative biology have unearthed a limited, ability for the mammalian heart to replenish lost CMs[2-5]. This exciting work has raised the possibility that innate growth mechanisms can be deployed for regeneration. Because innate regenerative responses are functionally absent in the adult mammalian heart, methods to improve the robustness of endogenous repair are needed for therapeutic heart regeneration to be realized.

The mechanisms for innate heart regeneration appear to be conserved across species. Following injury, pre-existing CMs proliferate to generate new CMs in zebrafish[6,7], newts[8,9], mice[4,10], rats[11], and pigs[12,13]. Accordingly, many groups are seeking to identify mitogens capable of promoting cardiomyogenesis. However, such work is challenging. Not only is the task of getting adult mammalian CMs to proliferate daunting but being able to identify rare proliferative events is difficult[1,14]. The challenge of identifying rare cycling CMs is compounded by the tendency of adult mammalian CMs to preferentially undergo endomitosis. For example, after injury to the mouse heart, almost 25% of CMs in the border zone re-enter the cell cycle, but only 3.2% of CMs divide[4]. Because most cycling CMs duplicate their genome but fail to undergo cytokinesis, simply assaying for an increase in the numbers of cycling CMs is ambiguous to bona fide cardiomyogenesis. Thus, assays for nucleoside incorporation by CMs or for the presence of proliferative markers on CMs may not entirely indicate regeneration. As more candidate factors for heart regeneration emerge, assays to better identify CM hyperplasia are needed.

Clonal analysis by lineage tracing is a valuable approach to assay for cardiomyogenesis because it allows for the direct visualization of cells and their progeny. Traditional approaches for clonal analysis involve rare labeling of single cells with a reporter gene. However, single-color lineage tracing of rare cells may be of limited value for infrequent events such as CM proliferation because the chances of labeling a proliferating CM are low[15]. Alternatively, multicolor lineage tracing can increase the sensitivity for clonal analysis[16]. Briefly, individual cells are genetically labeled with one of several fluorescent proteins at random, such that proliferating cells will generate homogeneously colored clusters of cells that can be resolved from neighboring fluorescent cells. This method has been used to trace growth across a variety of organs, and has been more recently applied to studies of mammalian heart regeneration[17,18]. While multicolor lineage tracing can detect clonal expansion of CMs in embryonic and neonatal stages, innate regenerative responses are not easily detected in the adult mouse heart after cardiac injury[17,19]. One approach to enhance the sensitivity of multicolor lineage tracing would be to increase the level of labeling and increase the probability for visualizing rare events. However, wider labeling comes at the cost of not being able to distinguish similarly labeled cells as rising from a common ancestor versus cells that were independently labeled with the same fluorophore. Presented

here is a protocol that uses nearest-neighbor modeling to identify clonally related CMs in the neonatal mouse heart. This method is unbiased, quantitative, and works over a range of labeling conditions.

PROTOCOL:

All procedures for handling mice, performing survival surgeries, and for harvesting hearts require approval by a local institutional animal use committee.

1. Mice for clonal analysis of CMs

1.1. Cross *Myh6-MerCreMer* mice[20] with *Gt(ROSA)26Sor*$^{tm1(CAG-EGFP,-mCerulean,-mOrange,-mCherry)Ilw}$ mice[21] to generate *Myh6-MerCreMer*; *R26R-Rainbow* bitransgenic mice.

1.1.1. Maintain stocks of mice homozygous for the *R26R-Rainbow* allele and cross with mice heterozygous for the *Myh6-MerCreMer* allele, such that all of the experiments are performed in mice heterozygous for the *R26R-Rainbow* allele. Alternatively, use mice homozygous for the *R26R-Rainbow* allele to increase color diversity, but this will require an imaging system that can faithfully discriminate EGFP, mCerulean, mOrange, and mCherry fluorophores.

1.2. Perform genotyping at the time of heart collection.

NOTE: *Myh6-MerCreMer* mice express a tamoxifen-inducible Cre recombinase in CMs and can be genotyped using the PCR primers CGTACTGACGGTGGGAGAAT (Cre-F) and GTGGCAGATGGCGCGGCAACA (Cre-R) to generate a 200 base pair (bp) product. *R26R-Rainbow* mice constitutively express *EGFP* but will stochastically express *mCerulean*, *mOrange*, or *mCherry* after Cre-mediated recombination. *R26R-Rainbow* can be genotyped with the PCR primers CTCTGCTGCCTCCTGGCTTCT (R26-F), CGAGGCGGATCACAAGCAATA (R26-R), and TCAATGGGCGGGGGTCGTT (*Rainbow*-R). The wild type allele gives a 350 bp product and the *Rainbow* allele gives a 250 bp product.

2. Cryoinjury and labeling of cardiomyocytes

2.1. Perform cryoinjuries as previously described with minor modifications[22,23]. Carefully place one-day-old mice (P1) on a bed of crushed ice for 3 min to induce cardiac arrest. Confirm adequate anesthesia by performing a toe pinch without reflexive withdrawal.

NOTE: Timing of anesthesia is critical. Survival can be impaired if mice are cooled for too long. Additionally, age of mice is critical. Regenerative capacity wanes sharply over the first week of life[10].

2.2. Position the anesthetized pup on a cold pack and under the objective of a surgical dissecting microscope. Use 6 mm microscissors to perform a thoracotomy between the 4th and 5th rib to expose the anterior surface of the heart.

2.3. Using a cryoprobe cooled in liquid nitrogen, place the probe on the left ventricle for 1-2 s. For sham injuries, perform thoracotomy without cryoinjury.

NOTE: The duration of cryoinjury dictates the extent of injury. Prolonged injuries will decrease survival rates.

2.5. Close the incision with 6-0 polypropylene suture attached to a 0.15 mm diameter needle.

2.6. Warm the pup quickly and manually stimulate to resume blood flow. Once the pup is spontaneously moving, place on a heating pad.

2.7. Inject recovered pups with 20 µg of tamoxifen intraperitoneally. At this point, inject analgesics, such as 0.25% bupivacaine along the suture line, if required. Return pups to the dam.

NOTE: Pups from a single litter are recovered together and returned to the dam together. This workflow improves survival and reacceptance of the litter by the dam.

3. Harvesting of hearts and processing for histological analysis

3.1. Harvest hearts at 20 days post injury (dpi), or postnatal day (P21), as regenerative responses can be observed by this time point[10]. First euthanize mice in a CO_2 inhalational chamber and then transfer to a surgical microscope.

NOTE: Euthanasia should be performed in compliance with institutional guidelines for animal use.

3.2. Once sufficient anesthesia is confirmed, perform a double thoracotomy using surgical microscissors to expose the heart. Then, perfuse the heart with 1 mL of 1 M KCl, 1 mL of 4% paraformaldehyde, and 1 mL of phosphate buffered saline (PBS).

NOTE: Adequate washing is needed to limit intraventricular blood clots that can cause autofluorescence during image acquisition.

3.3. After perfusion, dissect out the heart and place in a tube containing at least 1 mL of PBS.

3.4. Use a dissecting microscope to trim the hearts in order to optimize embedding. A horizontal slice below the atrium provides a good surface for embedding the hearts for short-axis sections.

NOTE: Alternatively, one can transversely cut the heart if long axis sections are desired.

3.5. Place trimmed hearts in 1 mL of 30% (wt/vol) sucrose at 4 °C for 12–18 h. Remove the samples from the sucrose solution. Embed trimmed tissue into blocks using tissue freezing media.

3.7. Freeze the embedded tissue quickly on dry ice and store blocks at −80 ˚C.

3.8. Section frozen whole-heart blocks into 10 µm sections with a cryostat, mount on charged slides, and store slides at −20 ˚C. Serially section hearts across 10 slides. This is done by mounting a section on each of the 10 slides before returning to the first slide to add another section. Such a strategy limits the chance of capturing the same CMs on multiple sections on a given slide.

NOTE: Be careful to limit exposure to light in order to prevent loss of fluorescence.

3.9. When ready for imaging, thaw slides and rehydrate in PBS. Add an anti-fade mounting medium before applying a #1.5 glass coverslip. Store cover-slipped slides in a covered container at 4 ˚C.

NOTE: For expected results for cryoinjury, the reader is directed towards prior work on this topic[23]. Cardiac sections have high levels of autofluorescence. Use of quenching agents, such as Sudan Black, prior to application of a coverslip can help to mitigate autofluorescence. However, caution should be used to avoid quenching fluorophores of interest.

4. Imaging

NOTE: The steps below apply to the use of a commercial microscope (see **Table of Materials**) that has an upright widefield fluorescence system with filter cubes equipped to discriminate mCerulean, EGFP, mOrange, and mCherry fluorophores, and software associated with this set up (see **Table of Materials**). Additional steps will vary depending on the exact imaging system that is used.

4.1. Image the entire slide at 5x magnification with only the GFP filter to create a slide map.

4.2. Select individual sections on the slide map that are intact and with limited autofluorescence for imaging. Due to the serial sectioning in sets of 10 slides, adjacent sections will be at least 100 µm apart, preventing imaging of the same CMs.

4.3. Perform tile scan imaging for the individual sections using a 20x objective with the mCerulean, mOrange, mCherry, and EGFP filters. Representative images are shown in **Figure 1**.

4.4. Save images in a format that allows for the individual channel information to be retrieved. Be careful to name the files in a format that will allow for identifying the heart, fluorophore, and position of the section within the heart. A systematic naming approach is required for the computational steps below.

5. Analysis of images to identify labeled CMs

5.1. Select the imaged sections that are intact from each set and distribute randomly to all screeners. Ensure that screeners adhere to the following steps for each image.

5.2. For each channel (mCerulean, mOrange, and mCherry), use the **'Color Balance'** tool on ImageJ[24] to adjust brightness and contrast of the image to help the user identify labeled cells.

NOTE: ImageJ is a free image analysis tool that can be downloaded from https://imagej.net. This protocol used version 2.0.0-rc-69/1.52i.

5.3. In ImageJ, select '**Set Measurements**' from the '**Analyze**' drop-down menu. In the window that opens, select '**Area | Center of mass | Bounding rectangle | Limit to threshold**'.

NOTE: The image subtraction module can sometimes be used to improve fluorophore signal relative to the background autofluorescence. However, if performed, subtraction should be done systematically for all images in a dataset.

5.4. With the image of interest open, select '**Tools**' and then select '**ROI manager**' from the '**Analyze**' drop-down menu. This will open, the '**ROI Manager**' window.

5.5. Select the '**Freehand selections**' tool from the ImageJ toolbar and trace the cell of interest (**Figure 2A**). Once traced, press '**t**' or use the '**Add [t]**' button in the '**ROI manager**' window (**Figure 2B**). Repeat for all the cells of a single channel.

NOTE: Segmentation and ROIs should be done for each channel separately.

5.6. Once all the cells for a given channel have been selected, save the ROI's as a .zip file. Do this by first selecting '**Deselect**' in the '**ROI Manager**' window to make sure no single ROI is highlighted. Then, click on the '**More**' tab in the '**ROI manager**' window and select the '**Save**' option. This will open the desktop navigator, where the ROIs are saved as a .zip file.

5.7. Export the measurements for each of the ROIs as a .csv file. To do this, make sure no single ROI is highlighted, again by selecting '**Deselect**', and then select '**Measure**' in the '**ROI manager**' window. This will open a new '**Results**' window with all the measurements selected earlier.

5.8. In the '**Results**' window select '**File**', then select '**Save as**' from the drop-down menu (**Figure 2C**). This will once again open the desktop navigator, so that the file can be saved in the .csv format.

NOTE: Make sure to save the ROIs and measurements before moving on to segmenting the next channel.

5.9. Be careful to name the files in a format that will allow for identifying the heart, channel, and position of the section within the heart. For example, 'HeartType#_Set#_Section#_Channel' (e.g., 'CI-1_Set4_S3_mOrange.zip' for the ROIs and 'CI-1_Set4_S3_mOrange.csv' for the corresponding measurements).

NOTE: The code below assumes the naming convention above. A different naming convention will require modification to the code.

6. Statistical analysis

NOTE: Cells carrying the same fluorophore are a mix of clonally derived cells (kin) and unrelated cells that underwent a random recombination event to express the same fluorophore (non-kin). Based on prior data[17], kin cells are assumed to have a closer physical proximity than non-kin cells expressing the same fluorophore. Thus, kin and non-kin cells can be differentiated based on a distance threshold. However, in order to determine the threshold, the nearest neighbor distributions for kin and non-kin cells need to be deconvolved. Fortunately, nearest neighbor distributions for non-kin cells can be estimated by evaluating the nearest neighbor values from each cell to the closest cell carrying a different fluorophore. Here, methodology is presented for statistically determining a threshold distance to define clonality and for assigning a probability of kinship between cells.

6.1. Using the .csv files obtained from section 5, calculate the nearest neighbor distances for each cell within a heart slice. Compile these results into within-channel (mCerulean-mCerulean, mOrange-mOrange, and mCherry-mCherry) and between-channel (mCerulean-mOrange, mOrange-mCherry, and mCherry-mCerulean) distances. Save the results into two .csv files: one for within-channel and one for between-channel distances.

6.1.1. Install the Python[25] code that requires the numpy and pandas libraries.

> pip install numpy; pip install pandas

6.1.2. Calculate distances and export .csv files for within-channel distances and between-channel distances using **python snippet_6_1.py <***path_to_ImageJ_results_directory***>.**

6.2. Read the within-channel and between-channel distances into an R[26] data frame, converting the heart identifiers into a categorical variable. Apply a threshold chosen ad-hoc by looking at nearest neighbor histograms (**Figures 3A,C**) to minimize outliers. Be sure to choose a threshold that covers over 90% of the points, while excluding the long tail to avoid overfitting to the noise at large nearest neighbor distances. This protocol uses a threshold of 400 μm (**Figure 3**). Use R script **Rscript snippet_6_2.R** to execute this step.

NOTE: Alternatively, one can model multiple slices as individual variables. This protocol limits the model complexity to the heart level for ease of modeling and estimation.

6.3. Obtain the natural logarithms for the within-channel and between-channel distances. Because cells are space-occupying, non-point entities, the minimum nearest neighbor distance distribution is centered around a non-zero value and has asymmetric tails since the maximum distance between pairs of cells can be substantially higher than the median values (**Figures 3A,**

C). Use the log-normal distribution to model such distributions (**Figures 3B,D**). Do this by using the R script **Rscript snippet_6_3.R**.

6.4. Use the following Bayesian model to estimate the means and variances of the kin (α_2, σ^2_2) and non-kin (α_1, σ^2_1) nearest-neighbor distributions, and the probability of within-channel distances arising from the kin distribution ($\pi=e^{\beta}/(1+e^{\beta})$):

$bw.dist_i \sim N(\alpha_1, \sigma^2_1 | l.lim<bw.dist_i<u.lim)$
$wi.dist_i \sim N(\alpha_1, \sigma^2_1 | l.lim<wi.dist_i<u.lim)*(1-\pi) + N(\alpha_2, \sigma^2_2 | l.lim<wi.dist_i<u.lim)*\pi$
$\sigma_1 \sim T_1(0,1|\sigma_1>0)$
$\sigma_2 \sim T_1(0,1|\sigma_2>0)$
$\alpha_1 \sim Unif(2,6)$
$\alpha_2 \sim Unif(1,5)$
$\beta \sim Unif(-5,0)$

where Unif(a,b) is the uniform distribution on the interval [a,b], $N(\mu,\sigma^2|l<x<u)$ is the normal distribution with mean μ and variance σ^2 constrained to the interval (l,u), $T_1(0,1|x>0)$ is the standard Cauchy distribution (T with 1 degree of freedom) constrained to the positive half-line, $bw.dist_i$ is the i^{th} logged between-channel pair distance and $wi.dist_i$ is the i^{th} logged within-channel pair distance. Note that the within-channel logged distances have a two-component mixture distribution reflecting the admixture of kin and non-kin cell pairs.

6.4.1. Provide the model with informative priors if possible. For example, this protocol expects the kin cells to be closer to each other than non-kin cells, so this is provided as a prior for the mean of the distribution of the distances. Then, use the runjags[27] and coda[28] R packages and the JAGS[29] program for parameter estimation (**Rscript snippet_6_4.R**).

6.5. Execute the JAGS model with the number of chains corresponding to the number of available CPUs. Use the **Rscript snippet_6_5.R** to perform this step. Allow the software to run until convergence.

6.6. Use metrics to test for convergence of the algorithm, such as effective sample size (available with runjags), the Gelman-Rubin convergence diagnostic (available with runjags), and its autocorrelation (autocorr/autocorr.diag available in coda). Verify the goodness of fit of the estimated model for the nearest neighbor distances by inspecting a Q-Q plot, which plots a quantile-level comparison of actual and estimated values for nearest neighbor distances using **Rscript snippet_6_6.R** (**Figure 4C,F**). This script requires the user to edit the values in lines 25-29 of snippet_6_6.R based on the output from Command 6.

NOTE: Ideally, an effective sample size (SSeff) of greater than 10,000 should be obtained for each of the parameters[30]. However, a SSeff greater than 2500 may be suitable for some complex models. The Gelman Rubin diagnostic is 1.00 for perfect convergence and must be less than 1.05 for satisfactory convergence. High autocorrelation indicates slow mixing within chains and is an

indicator of slow convergence. The autocorrelation plot must follow an exponential decay in case of a good model. The Q-Q plot of actual and modeled nearest-neighbor distributions (Figure 4C, F) should follow the diagonal as closely as possible.

6.7. After verifying the accuracy of the models, obtain the probability of a pair of cells being kin as a function of their distance. To determine a threshold distance for determining whether within-channel cells are likely to be kin, identify the distance at which the probability for being kin drops below 0.5. This can be done using **Rscript snippet_6_7.R.**

NOTE: Alternative thresholds can be chosen based on experimental goals. For some applications, thresholds may not be relevant and the probability of being a kin pair may be directly used to weight pairs of cells .

REPRESENTATIVE RESULTS:

Following the protocol for neonatal cryoinjury should yield P21 hearts with and without injury. Cryoinjured hearts have a well-circumscribed injury while the surface of sham hearts is smooth and homogeneous. In cryoinjured hearts, the area of injury should be consistent from heart to heart. After microscopy, images similar to **Figure 1** should be obtained. Note that the image resolution allows for identification of individual CMs and imaging conditions allow for each fluorophore to be resolved. Additionally, note that the injuries are not transmural as regenerative responses can be diminished with larger injuries[23,31]. After image segmentation, files with CM coordinates were obtained for each fluorophore for each section (**Figure 2C**). For the analyses presented, a total of 81 sections were used (46 from cryoinjured hearts and 35 from sham hearts). The sections for each heart varied from 9 to 17 based on heart size, and were distributed at least 100 μm apart, covering the entire heart. Vigilant attention to organization and naming of files cannot be overemphasized. Image segmentation provides the raw data needed for calculation of nearest neighbor distances and the modeling described in Step 6.

The histograms in **Figures 3A** and **Figure 3C** show nearest neighbor distances for within-channel and between channels pairs of CMs. While the histograms appear grossly similar, they differ at lower values, suggesting the presence of clustered kin cells in the within-channel distribution. The overall similarity of **Figures 3A** and **Figure 3C** can be explained by the rather small number of kin cells relative to non-kin cells, which are found in both within-channel and between-channel distributions. The log-normal distribution can be used to model within-channel and between-channel distributions (**Figures 3B,D**). To determine a threshold value for distinguishing kin and non-kin cells, the Bayesian model described in 6.4-6.7 was applied with 16 parallel chains to a representative heart during physiologic growth, yielding a kin threshold of 29.67 μm (**Figure 5**). This number is obtained by identifying the nearest neighbor distance at which the likelihood of a pair of cells being kin drops below 0.5. Importantly, this is the distance threshold for which adjacent cells carrying the same fluorophore have a greater probability of being clonally related than having arisen from independent recombination events in separate, un-related cells. Thus, it is possible to find kin pairs at distances greater than the threshold and non-kin pairs at distances less than the threshold. The distance estimates obtained from the model are listed in **Supplementary Table 1**. Representative results for nearest neighbor distance estimation are

shown in **Figure 4A,B,D,F**. Step 6.6 provides a list of diagnostics to estimate the correctness of the model. The effective sample size and the Gelman-Rubin Convergence Diagnostic are listed in **Supplementary Table 1** whereas the QQ plots are shown in **Figure 4C,F**. The autocorrelation plot obtained from Step 6.6 is shown in **Supplementary Figure 1**.

Using the presented methodology, results comparable to prior clonal analyses of CM expansion during physiologic growth and regeneration are attained. For example prior work has determined that clonally related CMs are smaller in size compared to non-clonal CMs[32]. Indeed, after applying the presented approach to 81 sections, clonally related CMs were found to be smaller than non-clonally related CMs (165.20 ± 13.79 μm^2 vs 252.37 ± 20.51 μm^2, p = 0.007, Welch t-test, **Figure 6A**). Moreover, the size of kin cells across experimental conditions can be assessed. Interestingly, kin cells had a non-significant trend towards being larger in injured hearts compared to sham hearts (189.29 ± 16.07 μm^2 vs 141.12 ± 10.57 μm^2, p = 0.076, Welch t-test, **Figure 6B**), perhaps relating to some element of concomitant hypertrophy of CMs following injury. Prior work has also shown that CM proliferation is not enriched, and may even decrease in the area of injury, following cryoinjury to the neonatal mouse heart[22,23,31]. Accordingly, there was no significant enrichment in the percentage of CMs that are clonally related between injured and sham hearts when looking at entire sections (8.57 ± 2.69 % in cryoinjured hearts vs 9.45 ± 1.23 % in sham hearts, p=0.788, Welch t-test, **Figure 6C**).

An important aspect of this protocol is the reproducibility of results under different labeling conditions that arise stochastically. In our experiments, we observed a large degree of variation in labeling across sham and cryoinjured hearts. However, as shown in **Figure 6D**, the threshold distance is consistent (Line of fit: y = -0.0011x + 38.30289, standard error of fit: 5.082, r^2=0.1911 - excluding the outlier) despite some hearts having more than a thousand cells labeled and other hearts being very sparsely labeled. The low r^2 value indicates that there is a weak correlation between the number of labeled cells and the threshold distance, and thus points to the strength of the model. Additionally, the percentage of labeled CMs that are clonally related is also consistent across hearts regardless of labeling efficiency (Line of fit: y = -0.00068x + 12.58379, standard error of fit: 1.538). Together, these data demonstrate the robustness of this protocol across experimental conditions.

FIGURE AND TABLE LEGENDS:

Figure 1: Multicolor labeling for clonal expansion. Tilescan image of cardiac sections from *Myh6-MerCreMer*; *R26R-Rainbow* mice after (**A**) sham injury with widespread labeling or (**B**) cryoinjury with limited labeling. Hearts carried fluorescent labels for mCerulean (blue), mCherry (red), and mOrange (green) CM, and EGFP (gray). The asterisk in panel B corresponds to the area of injury. Dashed boxes correspond to magnified insets. Arrows correspond to CMs that are clonally related based on their common fluorophore and close proximity. Such relationships are harder to assign by inspection alone for the more diffusely labeled heart in panel A. Scale bar is 500 μm.

Figure 2: Segmentation of CMs for clonal analysis. (**A**) Segmentation of mCherry CMs using the trace tool. This tracing technique captures original shape and location allowing for the possibility to use the segmented data for further analysis on cell shape and size. Scale bar is 100 μm. (**B**) ROI

manager window that records each segmented cell. (**C**) Results window that shows the saved measurements for each segmented cell including XY coordinates and the area.

Figure 3: Distributions of nearest neighbor distances. (**A**) Density histogram of the distances between a cell and its nearest neighbor from another channel. (**B**) Natural logarithm of the nearest neighbor between-channel distances plotted in (A). (**C**) Density histogram of the distances between a cell and its nearest neighbor from the same channel. (**D**) Natural logarithm of the nearest neighbor within-channel distances plotted in (C).

Figure 4: Modeled distribution of distances. (**A**) Density histogram of nearest neighbor kin distances. (**B**) Modeled distribution for nearest neighbor within-channel distances. (**C**) Q-Q plot showing goodness of fit of modeled within-channel distribution to actual data. (**D**) Density histogram of nearest neighbor non-kin distances. (**E**) Modeled distribution for nearest neighbor between-channel distances. (**F**) Q-Q plot showing goodness of fit of modeled between-channel distribution to actual data.

Figure 5: Estimated probability of kin pairs as a function of distance. The estimated probability of obtaining a kin pair as a function of nearest neighbor distance for within-channel pairs. The distance at which the probability drops below 0.5 (dashed line at 29.67 μm for this heart) is considered as the threshold to determine kin pairs. By allowing two components of the normal mixture model to have different variances, the likelihood of being a kin pair can artifactually appear to increase at higher nearest-neighbor distances (> 100 μm). Such artifact is a well-known feature of quadratic discriminant analysis models, and the user is encouraged to focus on the descending limb of the curve to determine a threshold for defining kin pairs.

Figure 6: Properties of kin and non-kin CMs. (**A**) Bar plot comparing mean cell area of kin cells (n = 1227 cells) with the entirety of the cells (n = 19733 cells). Error bars indicate SEM. (**B**) Normalized CM area for kin cells in injured (n = 579 cells) and sham hearts (n = 648 cells). Error bars indicate SEM. (**C**) Mean fraction of kin cells for injured (n = 3 mice) and sham hearts (n = 3 mice). (**D**) Kin threshold for different hearts (n = 6, 3 with injury, 3 with sham injury, 81 sections total) as a function of the number of cells imaged. The regression line is drawn excluding the outlier (marked by a plus). ** indicates $p < 0.01$ and NS indicates a non-significant difference.

Supplementary Figure 1: Autocorrelation plot for the model parameters. Autocorrelation plots for the means and standard deviations of the distributions and the likelihood parameter obtained from MCMC sampling. The exponential decay in the autocorrelation points to a good sampling and trustworthy results.

Supplementary Table 1: Estimates of parameters from Bayesian Modeling using JAGS. The subscripts 1 and 2 indicate non-kin and kin respectively. Alpha, sigma, and beta stand for the mean, the standard deviation, and the kinship odds parameter respectively. Columns of note are the mean value, Rhat – which shows the Gelman Rubin Convergence Diagnostic, and Neff – which shows the effective sample size.

DISCUSSION:

Multicolor lineage tracing is a powerful approach to identify patterns of organ growth with single cell resolution. However, a major limitation to multicolor lineage tracing is the need for sparse labeling of cells, which can reduce the sensitivity for identifying rare events. For organs like the heart with notoriously low levels of parenchymal cell turnover, this can lead to underestimates of growth responses. Presented here is a step-by-step protocol for performing clonal analysis of CM expansion during growth and regeneration in the neonatal mouse. Most importantly, an analytical framework is provided for identifying clones by modeling clonal probability with nearest neighbor distributions. This approach assumes that proliferating cells are in close proximity to each other and may not be applicable to systems where extensive migration of kin cells is expected.

This approach offers several advantages for clonal analyses. In comparison to previous multicolor lineage tracing studies[17], this protocol expands the sample size by imaging and segmenting entire heart sections, rather than analyzing only a subset of images per section. This method for segmentation also allows for ascertainment of physical features of CMs, including cell size and shape, providing a potential means to identify characteristics that can differentiate proliferating and non-proliferating cells. Finally, and most significantly, rather than assuming that clusters of cells carrying the same fluorophore are clonally related, the presented analysis provides an unbiased estimate of clonal probability that accounts for the density of cellular labeling.

Critical steps in the protocol include optimization of labeling efficiency, imaging, and segmentation. While this methodology works across a range of labeling conditions, extreme labeling conditions are likely to yield indeterminate results. Labeling that is too dense can limit the ability to deconvolve nearest neighbor distributions for kin and non-kin cells. By contrast, labeling that is too rare can limit the sensitivity for detecting rare events. Users are recommended to optimize labeling by testing a range of tamoxifen doses and ultimately selecting a dose that results in the largest amount of labeling while being able to deconvolve the kin and non-kin distributions. Sectioning and imaging conditions are also key to a successful experiment. High quality sections that enable a survey of anatomically distinct sets of cells are needed to avoid over-sampling of the same cells. Imaging conditions that result in a high signal-to-noise ratio for labeled cells is critical as systematic under-representation of fluorophores can bias estimates of non-kin nearest neighbor distances. Particular attention to the mCerulean fluorophore is needed as this fluorophore is especially sensitive to photobleaching and loss of signal. Finally, careful segmentation is needed because this method relies on accurate assessments of intercellular distances. When multiple users are performing segmentation, users should be issued images for segmentation that are randomized across treatment groups in order to minimize user bias.

Several aspects of this approach have the potential to be modified to address limitations. For example, after tissue harvesting and processing, the major bottleneck is image segmentation. With advances in machine learning and computer vision, automated methods for image segmentation are increasingly accurate and can increase the throughput of image segmentation. Because a file of cellular coordinates is all that is needed for this approach, the code provided is

compatible with object identification using automated methods. Similarly, while this method is currently configured for 2-dimensional datasets, 3-dimensional datasets are likely to be more widely utilized with improvements in tissue clearing methodology and light sheet microscopy. This approach can easily accommodate such datasets by simply adjusting the code to compute 3-dimensional distance measures. Overall, future iterations of this approach are expected to be able to accommodate large datasets that encompass whole organs.

In summary, this protocol generates a pipeline for imaging and analyzing multicolor data that is scalable to large numbers of images. Prior work using multicolor lineage tracing for detecting cardiomyogenesis has relied on small scale assessment of samples for clonal analysis. The presented methodology allows for rapid and rigorous assessments of regenerative therapeutics using large data sets.

ACKNOWLEDGMENTS:

This work was funded by an R03 HL144812 (RK), a Duke University Strong Start Physician Scientist Award (RK), a Mandel Foundation Seed Grant (RK), and a T32 HL007101 Training grant (DCC). We would additionally like to acknowledge Evelyn McCullough for assistance with mouse husbandry and Dr. Douglas Marchuk and Matthew Detter for helpful comments and discussion. Finally, we would like to thank Purushothama Rao Tata for kindly providing *R26R-Rainbow* mice.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES

1. Karra, R., Poss, K. D. Redirecting cardiac growth mechanisms for therapeutic regeneration. *Journal of Clinical Investigation.* **127** (2), 427-436 (2017).
2. Bergmann, O. et al. Dynamics of Cell Generation and Turnover in the Human Heart. *Cell.* **161** (7), 1566-1575 (2015).
3. Bergmann, O. et al. Evidence for cardiomyocyte renewal in humans. *Science.* **324** (5923), 98-102 (2009).
4. Senyo, S. E. et al. Mammalian heart renewal by pre-existing cardiomyocytes. *Nature.* **493** (7432), 433-436 (2013).
5. Mollova, M. *et al.* Cardiomyocyte proliferation contributes to heart growth in young humans. *Proceedings of the National Academy of Sciences U.S.A.* **110** (4), 1446-1451 (2013).
6. Kikuchi, K. et al. Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes. *Nature.* **464** (7288), 601-605 (2010).
7. Jopling, C. et al. Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature.* **464** (7288), 606-609 (2010).
8. Oberpriller, J. O., Oberpriller, J. C. Response of the adult newt ventricle to injury. *Journal of Experimental Zoology.* **187** (2), 249-253 (1974).
9. Oberpriller, J., Oberpriller, J. C. Mitosis in adult newt ventricle. *Journal of Cell Biology.* **49** (2), 560-563 (1971).

10. Porrello, E. R. et al. Transient regenerative potential of the neonatal mouse heart. *Science.* **331** (6020), 1078-1080 (2011).
11. Wang, H. et al. Natural Heart Regeneration in a Neonatal Rat Myocardial Infarction Model. *Cells.* **9** (1), (2020).
12. Ye, L. et al. Early Regenerative Capacity in the Porcine Heart. *Circulation.* **138** (24), 2798-2808 (2018).
13. Zhu, W. et al. Regenerative Potential of Neonatal Porcine Hearts. *Circulation.* **138** (24), 2809-2816 (2018).
14. Kadow, Z. A., Martin, J. F. Distinguishing Cardiomyocyte Division From Binucleation. *Circulation Research.* **123** (9), 1012-1014 (2018).
15. Roy, E., Neufeld, Z., Livet, J., Khosrotehrani, K. Concise review: understanding clonal dynamics in homeostasis and injury through multicolor lineage tracing. *Stem Cells.* **32** (12), 3046-3054 (2014).
16. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature.* **450** (7166), 56-62 (2007).
17. Sereti, K.-I. et al. Analysis of cardiomyocyte clonal expansion during mouse heart development and injury. *Nature Communications.* **9** (1), 1-13 (2018).
18. Xiao, Q. et al. A p53-based genetic tracing system to follow postnatal cardiomyocyte expansion in heart regeneration. *Development.* **144** (4), 580-589 (2017).
19. Ali, S. R. et al. Existing cardiomyocytes generate cardiomyocytes at a low rate after birth in mice. *Proceedings of the National Academy of Sciences U.S.A.* **111** (24), 8850-8855 (2014).
20. Sohal, D. S. et al. Temporally regulated and tissue-specific gene manipulations in the adult and embryonic heart using a tamoxifen-inducible Cre protein. *Circulation Research.* **89** (1), 20-25 (2001).
21. Red-Horse, K., Ueno, H., Weissman, I. L., Krasnow, M. A. Coronary arteries form by developmental reprogramming of venous cells. *Nature.* **464** (7288), 549-553 (2010).
22. Polizzotti, B. D. et al. Neuregulin stimulation of cardiomyocyte regeneration in mice and human myocardium reveals a therapeutic window. *Science Translational Medicine.* **7** (281), 281ra245 (2015).
23. Polizzotti, B. D., Ganapathy, B., Haubner, B. J., Penninger, J. M., Kuhn, B. A cryoinjury model in neonatal mice for cardiac translational and regeneration research. *Nature Protocols.* **11** (3), 542-552 (2016).
24. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods.* **9** (7), 676-682 (2012).
25. van Rossum, G. Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam. (1995).
26. Team, R. C. R: A language and environment for statistical computing. (2013).
27. Denwood, M. J. runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software.* **71** (9), 1-25 (2016).
28. Plummer, M., Best, N., Cowles, K., Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R News.* **6** (1), 7-11 (2006).
29. Plummer, M. in *Proceedings of the 3rd international workshop on distributed statistical computing.* Vienna, Austria. 1-10 (2003).

30. Kruschke, J. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press. (2014).
31. Darehzereshki, A. et al. Differential regenerative capacity of neonatal mouse hearts after cryoinjury. *Developmental Biology.* **399** (1), 91-99 (2015).
32. Kimura, W. et al. Hypoxia fate mapping identifies cycling cardiomyocytes in the adult heart. *Nature.* **523** (7559), 226-230 (2015).

Figure 1

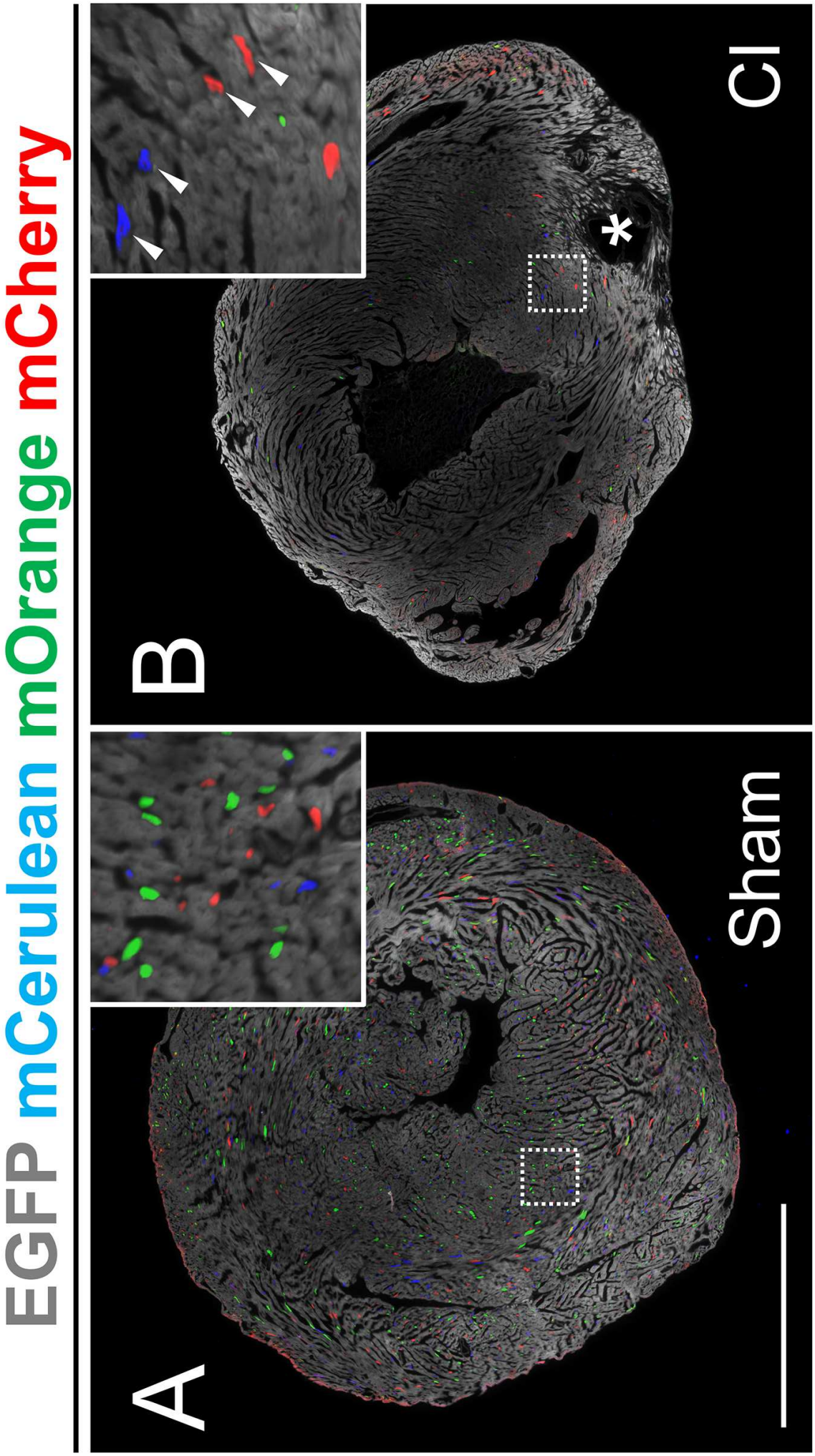
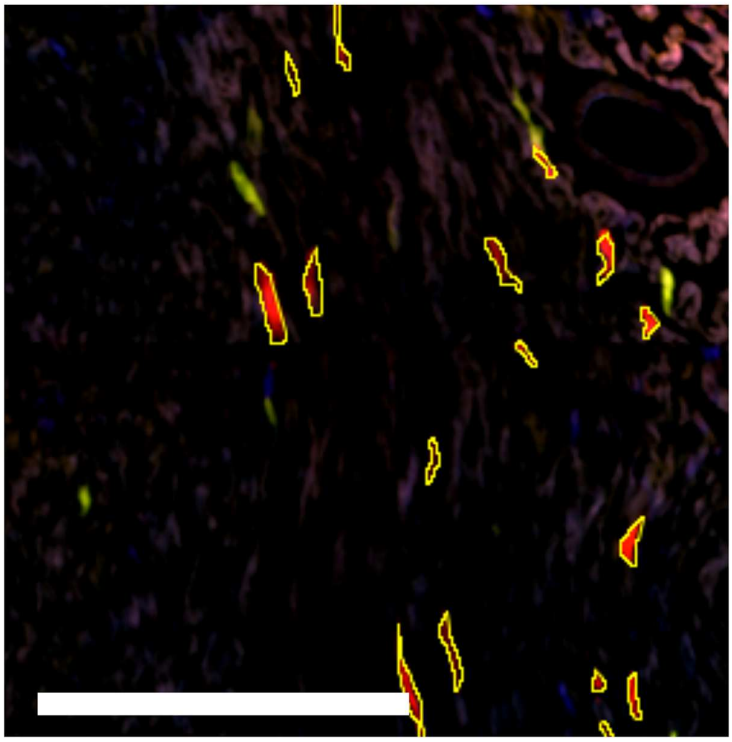


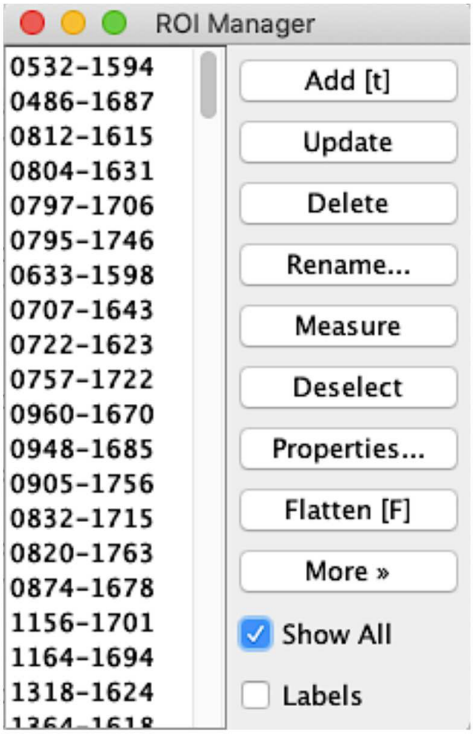
Figure 2

Click here to access/download;Figure;Figure 2.eps

A



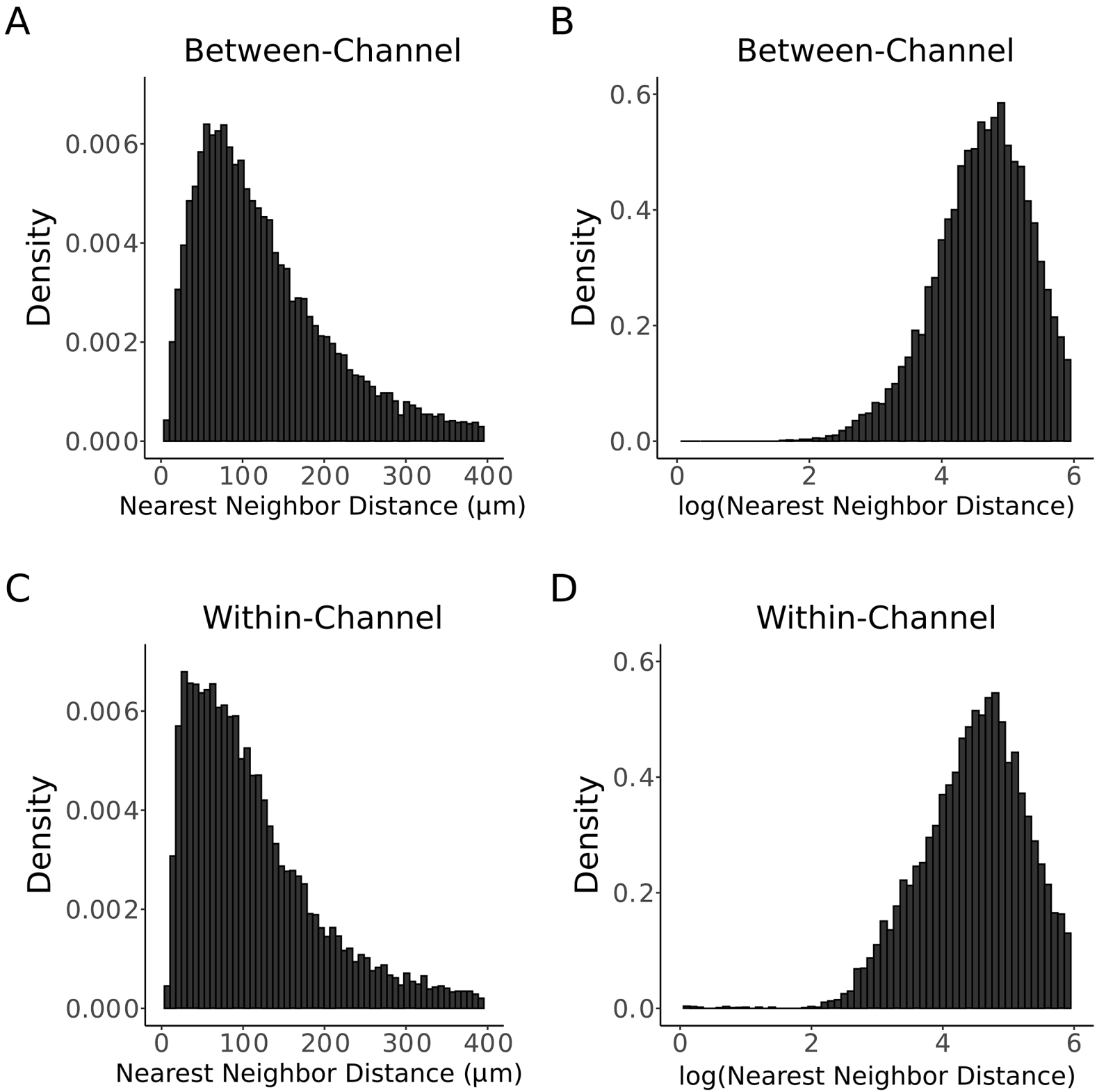
B



C

Results

	Area	XM	YM	BX	BY	Width	Height
1	100.172	1085.224	363.052	1078.024	347.991	14.982	28.602
2	23.652	1148.597	330.659	1146.805	327.561	4.086	6.810
3	82.086	1099.938	553.606	1095.730	542.757	8.172	21.111
4	54.724	1111.004	548.973	1108.669	538.671	4.767	17.706
5	44.985	1160.979	541.386	1157.020	535.947	9.534	14.301
6	34.318	1189.462	539.711	1186.984	534.585	4.086	13.620
7	22.261	1088.142	431.377	1086.196	425.625	4.086	10.896
8	29.681	1118.365	482.269	1116.841	471.933	4.086	19.068
9	17.623	1105.761	492.158	1103.902	486.234	3.405	11.577
10	15.768	1173.071	516.155	1169.959	511.431	6.129	8.172
11	57.043	1137.143	656.898	1133.866	638.778	6.810	30.645
12	42.202	1147.861	645.343	1144.762	635.373	6.129	21.111
13	38.492	1195.782	617.118	1193.113	610.176	6.129	12.939



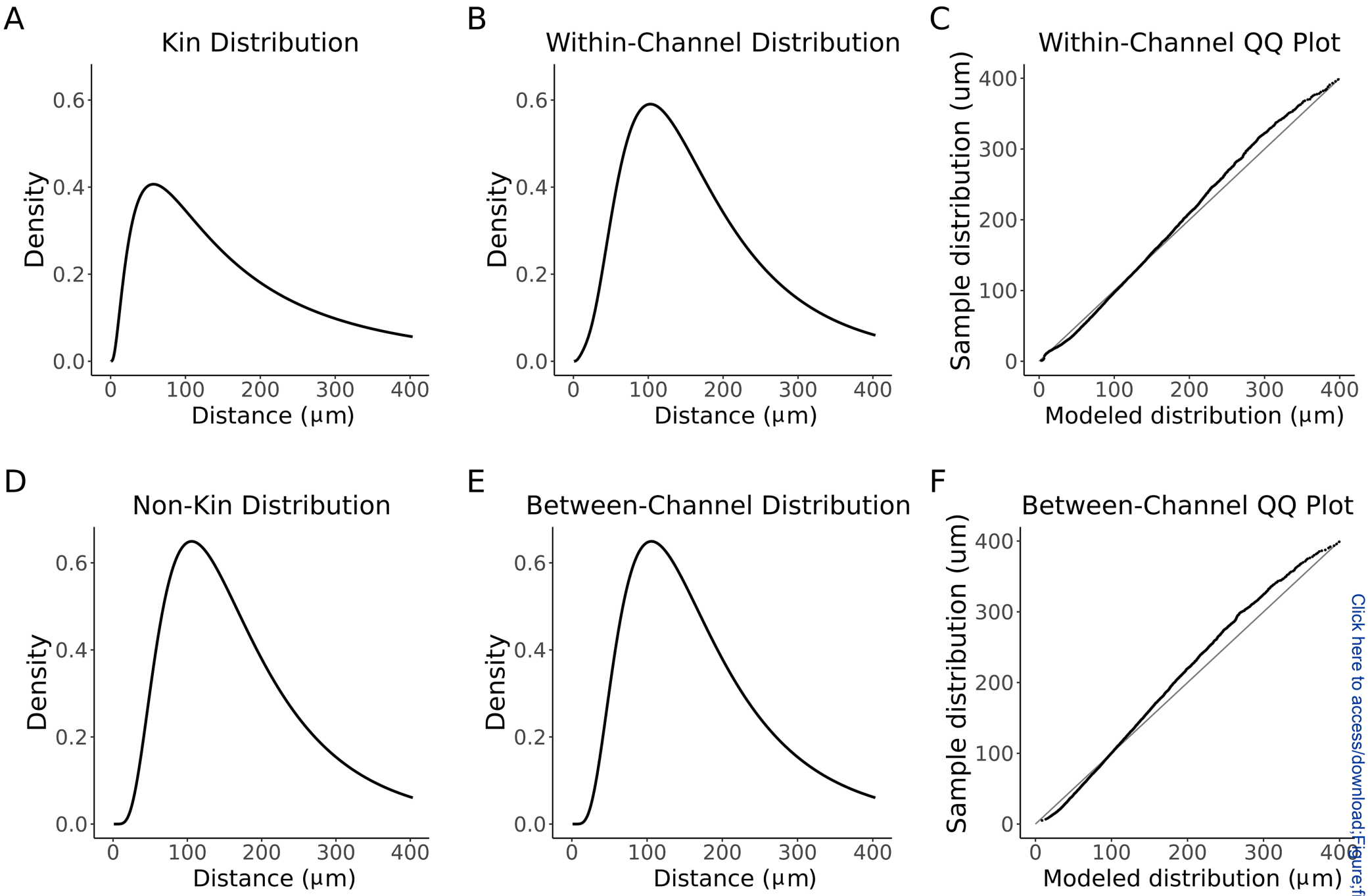
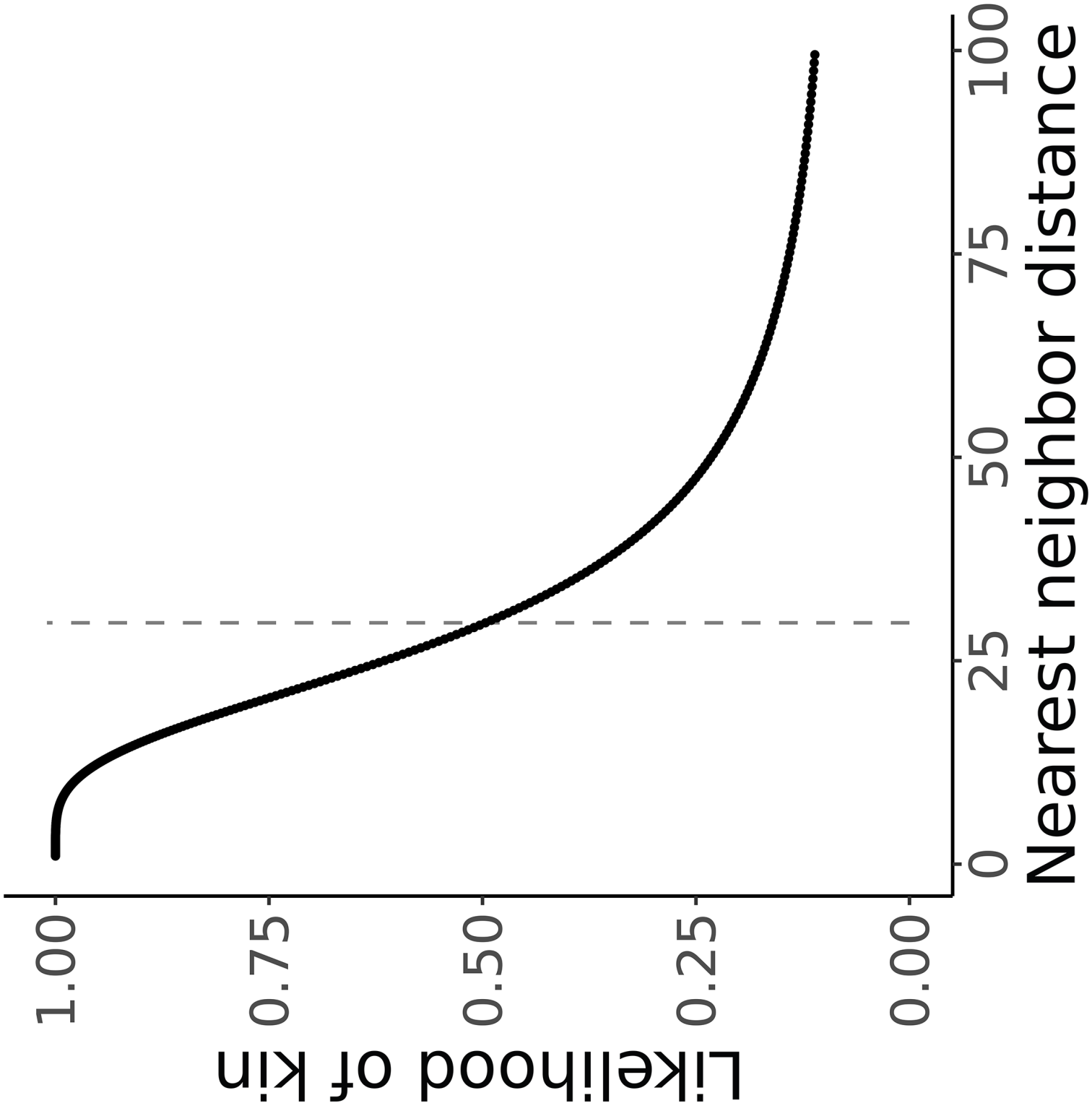
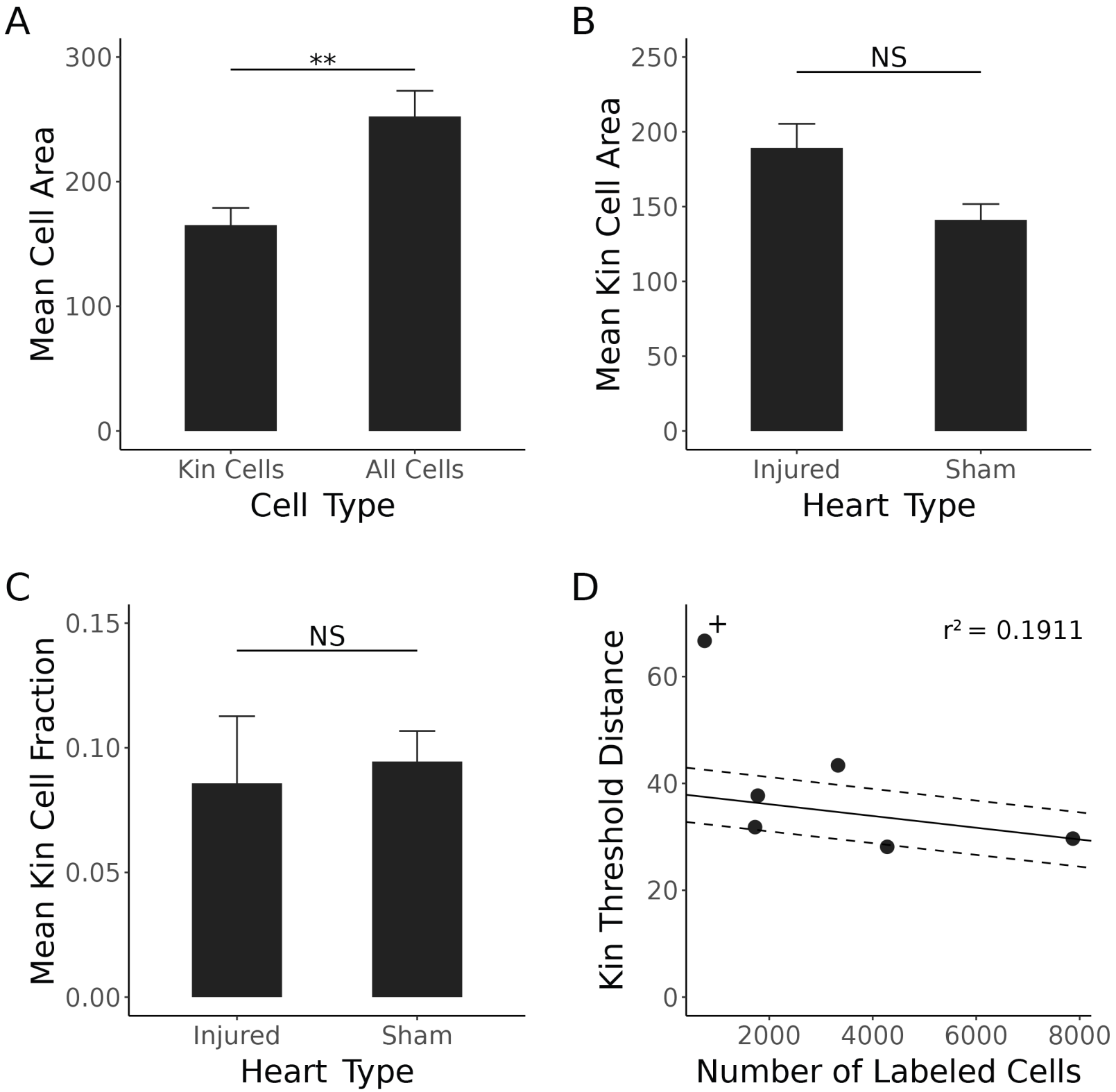


Figure 5

Click here to access/download;Figure;fig5.eps





Name	Company
#1.5 glass coverslip	FisherScientific
6-O prolene	Ethicon
Anti-fade mounting medium	FisherScientific
CO2 inhalational chamber	
Cold pack	
Cryomolds	VWR
Cryoprobe	World Precision Instruments
Filter cubes	
Gt(ROSA)26Sor$^{tm1(CAG-EGFP,-mCerulean,-mOrange,-mCherry)Ilw}$ mice	
ImageJ software	
KCl 1M	FisherScientific
Leica CM3050 cryostat	
Liquid Nitrogen	
Microscissors, 6mm	World Precision Instruments
Myh6 -CreERT2 mice	The Jackson Laboratory
Needler holder	World Precision Instruments
Paraformaldehyde 4%	FisherScientific
Phosphate buffered saline	
Python	
R	
Rotating Shaker	
Stereoscope	
Sucrose 30% (wt/vol)	FisherScientific
Surgical dissecting scissors	World Precision Instruments
Syringe for tamoxifen	VWR
Tamoxifen, 20 µg	Sigma
Tissue Freezing Media	VWR
White Frosted/Plus slides	Globe Scientific
Zeiss Axio Imager M1 upright widefield fluorescence system	
Zen 2.5 Blue software	

Catalog Number	Comments
12-544E	
8706H	
00-4958-02	
15160-215	
501313	
	https://imagej.net
LC187951	
14003	
005657	
14109	
AC416785000	
	https://www.python.org/
	https://cran.r-project.org/
BP220	
14393	
BD328438	
T5648	
15148-031	
1358W	

RESPONSE TO EDITORIAL COMMENTS:

1. The editor has formatted the manuscript to match the journal's style. Please retain and use the attached version for revision.

Response: We have retained these revisions.

2. Please address specific comments marked in the manuscript.

Response: We have addressed each comment and provided specific responses in the revised manuscript.

3. Please ensure that the length of the protocol section is no more than 10 pages.

Response: We have moved the code snippets to supplemental files to make the protocol section less than 10 pages.

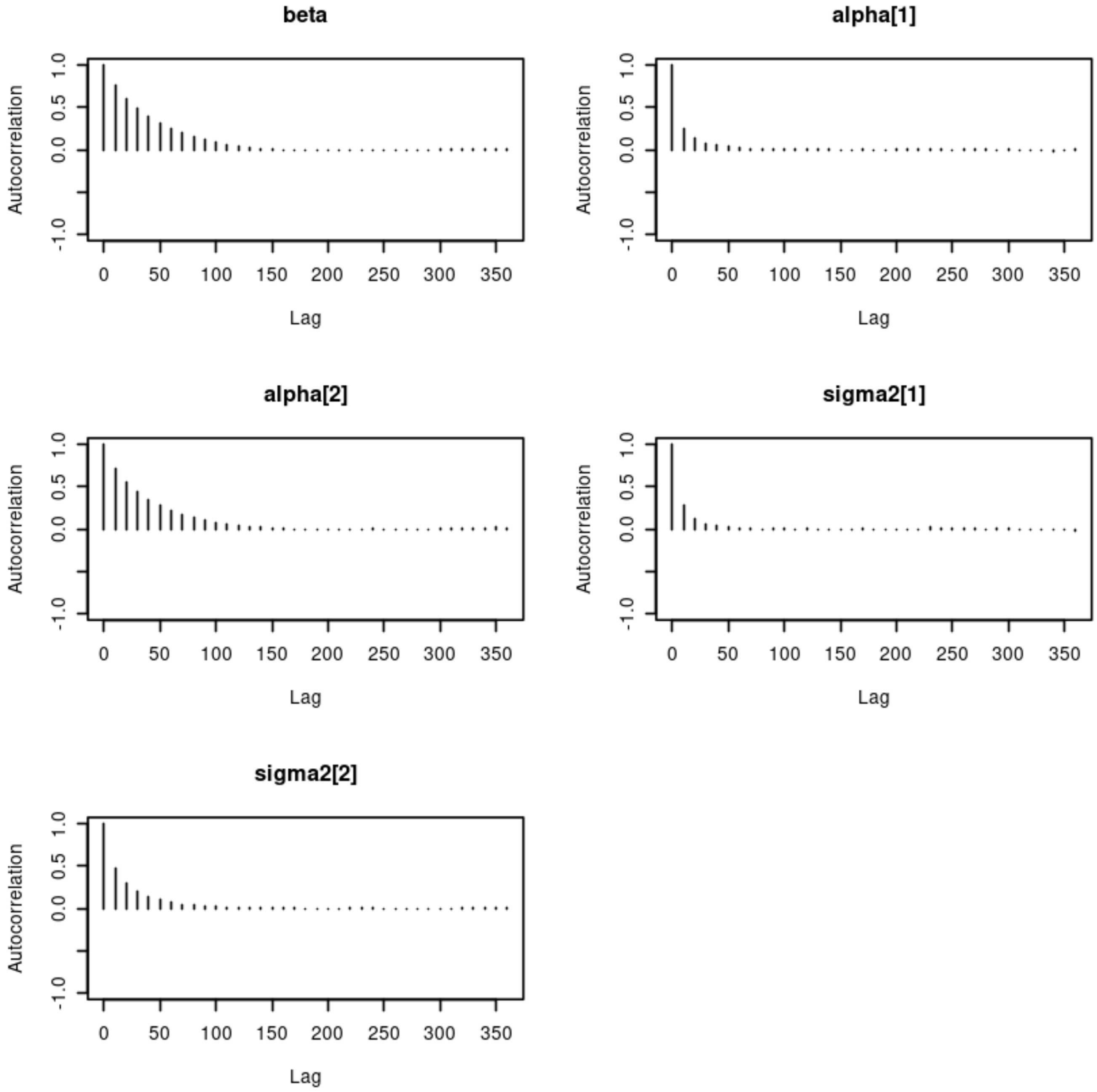
4. Please ensure that the highlighted section is no more than 2.75 pages including headings and spacings. Also if a step is highlighted coded/script etc present in the manuscript which shows how to do the step needs to be highlighted as well. In this case, please move the code to supplementary files and then specify the terminal line commands used to execute the code/script instead.

Response: Thank you. We have made the recommended changes.

5. Once done please proofread the manuscript well.

Supplementary Figure 1

Click here to access/download;Supplemental File (Figures, Permissions, etc.);Supplementary Figure 1 .eps



Supplementary Table 1

	mean	sd	2.50%	25%	50%	75%
beta	-1.4613754	0.17800387	-1.8118517	-1.58	-1.46104	-1.342505
alpha[1]	4.661229	0.00717198	4.64719	4.65645	4.66122	4.66603
alpha[2]	4.0487038	0.07703818	3.88984	3.998477	4.05098	4.10202
sigma2[1]	0.6143533	0.00768985	0.5994909	0.609163	0.614265	0.61949
sigma2[2]	0.9810601	0.05423567	0.8831419	0.943655	0.97796	1.01513

97.50%	Rhat	n.eff
-1.111318	1.002996	4100
4.67531	1.001426	16000
4.1919	1.002448	5500
0.629666	1.001046	58000
1.096141	1.001116	39000

Click here to access/download

Supplemental Coding Files

code_snippets.zip