**Single Cell Phenotyping and Survival Analysis.**

By Derek H. Oakley MD, PhD. Massachusetts General Hospital & Harvard Medical School.  
Mail: doakley at partners dot org.  
  
Please Cite: Oakley DH, Klickstein N, Commins C, Chung M, Dujardin S, Bennett RE, Hyman BT, Frosch MP. Continuous Monitoring of Tau-Induced Neurotoxicity in Patient-Derived iPSC-Neurons. J Neurosci. 2021 May 12;41(19):4335-4348. doi: 10.1523/JNEUROSCI.2590-20.2021. Epub 2021 Apr 23. PMID: 33893219; PMCID: PMC8143197.

Written on Fiji ImageJ 1.52e.  
  
**Overview:**  
The Single Cell Phenotyping and Survival Analysis toolset facilitates stereologic counting and tracking of individual cells in time-lapse images and scoring of user-defined primary and secondary morphologic phenotypes as well as cell death. Survival analysis allows censoring of individual cells that are lost or become obscured. Designed to work with imaging experiments performed in standard multiwell plates, up to 96 wells. Contains built-in Z-stack batch processor pre-set for image naming from Biotek Cytation5 instruments, but user-adjustable to other file types.  
  
Outputs include: Survival probability curves optionally stratified by primary or secondary features, cumulative event plots for primary and secondary features, and montaged Z-stacks of individually tracked cells, sorted by user-defined criteria. Annotated and flattened images of each counted Z-stack are also produced. Statistical comparisons are performed for survival or cumulative event analyses of two groups using the log-rank test and p values are reported in 0.05, 0.01, 0.001, and 0.0001 increments based on the Chi-squared distribution. Multiple pairwise-comparisons can subsequently be displayed on the same plot. Additional features include combination of single cell data from multiple replicates with an option to change well designations and resampling of data to change the first timepoint, allowing improved group-assignment for survival analysis.  
  
**Notes:**  
- Requires 'Action Bar' plugin for toolbar functionality and contains Cycloid Grid Macro for stereology grid application (http://imagejdocu.tudor.lu/doku.php?id=plugin:utilities:action\_bar:start & https://imagej.nih.gov/ij/macros/Cycloid\_grid.txt)  
- Must be installed in the ImageJ/macros/toolsets folder.  
- Designed for image names starting with standard well name and ending with a three digit frame number (i.e.B9\_...\_001.tif).  
- Keyboard shortcuts [x] are used to call individual macros.  
- Cell tracking is done with a custom ImageJ tool. Phenotypes are set with Cell State [a] macro.  
- Cell State macro allows cell death and location tracking, censoring of lost cells, and assignment of up to two features per timepoint.  
- Results are saved in subfolders created in the user-designated output folder.  
- Automatic output during cell tracking includes .csv files of individual cell location and phenotype for every timepoint counted, overlays of counted cells and cycloid grid if used, and text files summarizing data from each individual cell counted.  
- Single cell text data files are strings in the format described below and can be read into R for confirmatory survival and event analysis using the R survminer package or other software.  
- Survival and event analysis is performed in a guided fashion from the toolbar (see below).  
- Additional outputs generated from the data analysis portion of the toolbar include Kaplan-Meier survival probability curves comparing user-defined groups, text files describing group survival at each timepoint, aggregated group statistics, and extracted/montaged single cells.  
- Kaplan-Meier survival probability and log-rank calculations are performed as described in Bewick et. al Critical care 2004, 8:389-394.  
  
**Installation:**  
- Place macro file in the '/macros/toolsets' directory of FIJI or ImageJ and restart ImageJ.  
- Select '>>' on the right side of the ImageJ toolbar and then click on the macro name to install.  
- The macro and toolbar should then auto-run. If not, press [t] to call toolbar.  
- A cross-shaped cell tracking tool (Survival Tool) will install in the ImageJ toolbar and is used in conjunction with the 'Cell State [a]' macro to set and track cell states.  
- Optional: Run the first part of the Setup macro to create a ring-shaped cursor for improved visibility of small objects during tracking. Creates a file called 'crosshair-cursor.gif' in the '/images' directory of FIJI or ImageJ. Requires ImageJ restart to take effect. Deleting the 'crosshair-cursor.gif' file will revert ImageJ back to the default cursor.  
  
**Standard order of operations:**  
1. Prepare Images:  
-Run Setup macro  
-Align Z-stacks (optional)  
-Merge multichannel images (optional)  
2. Data Aqcuisition:  
-Open Z-Stack [o]  
-Apply Cycloid Grid for unbiased counting [c] (optional)  
-Set Cell State of first cell to count [a]  
-Click through Z-stack, following a single cell with the Survival Tool (+) and update cell state with [a] as necessary.  
3. Data Analysis:  
-If desired, resample data to change first timepoint, combine replicates, rename wells, or extract single cell images.  
-Calculate results and plot survival curves and/or cumulative events [g].  
  
**Toolbar Functions:**  
-Toolbar menu can be called with [t] keyboard shortcut.  
  
Instructions: Calls this HTML window.  
  
**1. Prepare Images:**   
Setup: Must be run prior to tracking cells. Prompts user to name the project (for folder and file naming purposes), designate if starting with Z-stacks or unstacked frames, and designate if using Biotek Cytation5 or other image naming format. Input and output directories are also selected.  
  
Change Working Dir: Allows user to alter the working directories used to find files during data acquisition and analysis. Can be useful if moving files or troubleshooting errors.  
  
Z-Stack Align: Allows user to stack single frames into Z-stacks and optionally perform SIFT alignment of images. Also permits left- or right truncation of Z-stack time series (to begin or end the experiment not at the first or last frame of acquired data)  
  
Merge Channels: Merges multi-wavelength image stacks based on the detected wavelengths present.  
  
**2. Data Acquisition:**   
Open Z-Stack [o]: Opens next Z-stack in list. Requires user to select a well. Image counter in menu displays the overall file number of the image in the folder and can be re-set to zero to move backwards (i.e. from row D to row A or column 12 to column 1)  
  
Manual Set Well [m]: Allows user to manually set well metadata for an image. Useful if Z-stack opener has an error or using images with non-standard names.  
  
Cycloid Grid: Applies the cycloid grid macro for stereology. Optional. Default settings can be overridden to use full macro (https://imagej.nih.gov/ij/macros/Cycloid\_grid.txt)  
  
Symbol Key [k]: Generates a key for the symbols used on flattened images to track the primary feature in counted cells. Reports cell classes as 'initial state'+'endstate'+'endlife'. In positions 1 & 2: A=Primary Feature, O=No Primary Feature, In position 3: A=Alive, D=Dead (at last timepoint counted). Different symbols reflect different cell classes/censoring as indicated in the key.  
  
Cell State [a]: Allows the user to iteratively change the cell state with respect to: Life, Primary Feature, Secondary Feature, and Censoring. Also allows deletion of cells with errors and movement on to the next image for tracking. There is also an option to allow use on non-Z-stacks for counting of single timepoints.  
  
**3. Data Analysis:**   
Resample First Timepoint: Allows left truncation of experiments and re-assigns cell fates at the newly designated t0. Can be helpful when using a primary or secondary feature occurring before a certain timepoint to define groups for survival analysis because cells must survive until the development of the feature, which biases the beginning of survival curves.  
  
Combine Replicates or Rename Wells: Allows combination of single cell text files from multiple replicates or wells by renaming the well names and files and moving into a common folder for later analysis.  
  
Extract and Montage Cells: Uses Results/Location .CSV files and Z-stacks to extract centered time-lapse Z-stacks of each tracked cell. Individual tracked cell images are placed into a folder from which they are then called for montage creation. Montages of user-defined size can be sorted by Cell Survival, Primary Feature, Secondary Feature, or left unsorted. There are two ways to montage cells from specific wells. 1. Place corresponding locations files in a separate folder, then change the .csv folder location with the Change Working Dir button and run the macro. 2. Alternatively, well-specific single cell images can be placed in a distinct folder prior to montage generation. It is also possible to extract tracked xyz positions from images with file names that do not match those on which tracking was initially performed (i.e. single wavelengths or brightfield images). To do this, replace the 'ImageName' columns in the .CSV location files with the desired file names using find/replace. Concatenating .CSV files prior to name replacement will speed the process (in windows cmd.exe from Locations folder: copy \*.csv AllWells.csv), then remove extra column headings, move new file into empty directory and run extraction macro).  
  
Survival and Event Analysis: Uses single cell summary data text files to create Kaplan-Meier survival probability curves or Cumulative Event graphs comparing one or multiple groups. Allows comparison of cell death in the following categories: Within wells (subset by Primary or Secondary Feature), Between wells (with option to subset by Primary or Secondary Feature), For all cells in the single cell data folder (subset by Primary or Secondary Feature), or Cumulative Event Analysis comparing the occurrence of the Primary or Secondary Feature between two groups of wells. Statistical calculations are performed as noted above for comparisons of two groups. All plots have customizable legends and axis titles. Plots with two curves have user specified line colors. Plots with more than two curves have the option of several color palettes.  
  
Note: Toolbar functionality is dependent on maintaining both macro file name and file location in '/macros/toolsets'  
  
**List of data folders created in output directory:**  
Useful folders:  
\_Aligned\_Z\_Stacks\: Contains aligned and stacked images from single frames. Will contain unaligned Z-stacks if SIFT alignment is not used. Contains channels as separate images.  
\_Merged\_Z\_Stacks\: Contains merged and aligned Z-stacks. This is the folder that is used to open color Z-stacks for counting.  
\_Raw\_Data\_Single\_Cells\: Contains individual cell .txt files. These files are saved after every cell and contain summary data for each tracked cell. These are the data files used by the graphing macro and are also used for importing data into R (see below).  
\_CSV\_Location\_Files\: Contains .CSV files with location and cell phenotype information from every cell tracked at every timepoint. Separate file per image. CSV files also contain data from deleted cells - indicated by the delete column.  
\_Aggregate\_Group\_Survival\_Plots\: Contains plot images created by grapher. Autosaved when plots are produced.  
\_Probability\_Group\_Survival\: Contains the survival probabilities and cumulative event probabilities.  
\_Aggregate\_Group\_Lifespan\_Statistics\: Contains statistics about each group graphed in the graphing macro.  
  
Less useful, but needed for data analysis:  
\_Raw\_Group\_Survival\: Contains raw survival and cumulative event data.  
\_Aggregate\_Group\_Lifespan\: Contains raw lifespan data from graphing macro.  
\_Aggregate\_Group\_Histogram\: Used in survival and cumulative event analysis calculations.  
\_Aggregate\_Group\_Censor\_Lifespan\: Contains the lifespans of censored objects, used in accounting for censoring in data analysis and graphing.  
\_Aggregate\_Group\_Censor\_Histogram\: Used in survival and cumulative event analysis calculations.  
  
**Lookup for string identities in single cell .txt data:**  
[x] designates the array element number needed to call the specific string feature after splitting string by ';' (arrays in ImageJ begin with 0).  
Notes: Integer values below default to 9999 if unfilled by data. '9999' is subsequently used to indicate the absence of an event.  
This program was initially written to track protein aggregate formation in neurons as a primary feature, so variables referring to primary feature often refer to an 'aggregate'.  
  
String identities in single cell .txt data (\*\*\*\*=most useful fields for data analysis in R or other program):  
[0]CellClass - A string describing the state of the cell with respect to the primary feature. Reports cell class as 'initial state'+'endstate'+'endlife'. In positions 1 & 2: A=Primary Feature, O=No Primary Feature, In position 3: A=Alive, D=Dead (at last timepoint counted)  
[1]well+-+position+-+CellName+ (CellName2) - Unique cell identifier (per experiment, will be non-unique across replicates)  
[2]lifespan - lifespan of cell max is = timepoints counted\*\*\*\*  
[3]xorig - the x position of the cell in the first counted frame  
[4]yorig - the y position of the cell in the first counted frame  
[5]aggregatelifespan; - lifespan of the primary feature\*\*\*\*  
[6]aggstartx - the x position of the cell when the primary feature occurs - an asumption is made that cells to not enter, leave, and then re-enter the primary feature state. True for below as well.  
[7]aggstarty - the y position of the cell when the primary feature occurs  
[8]aggstartz - the z position of the cell when the primary feature occurs - this is equivalent to timepoint.\*\*\*\*  
[9]feature2lifespan - lifespan of the secondary feature\*\*\*\*  
[10]f2startx - the x position of the cell when the secondary feature occurs  
[11]f2starty - the y position of the cell when the secondary feature occurs  
[12]f2startz - the z position of the cell when the secondary feature occurs - this is equivalent to timepoint.\*\*\*\*  
[13]well - Well that the cell is located in (i.e. A1, C5)\*\*\*\*  
[14]name -Image name  
[15]project - User specified project name  
[16]initialstate - First element in cell class above. The primary feature state of the cell at time 0: A=Primary Feature, O=No Primary Feature  
[17]endstate - Second element in cell class above. The primary feature state of the cell at the endpoint: A=Primary Feature, O=No Primary Feature  
[18]endlife - the Alive/Dead state of the cell at the endpoint: A=Alive D=Dead  
[19]timepoints - number of timepoints (Z-planes) in the experiment - used for calculations of survival and censoring\*\*\*\*  
[20]censor - whether or not the cell is 'censored' vs 'uncensored'\*\*\*\*  
  
The following lines of code can be used to **import a folder of single cell .txt files into R** from the folder 'datapath' for analysis with Survminer or other package:  
Replace single quotes with double quotes below prior to running.  
  
setwd(datapath)  
filelist <- list.files(path = datapath, pattern = '\*.txt')  
datalist <- lapply(filelist, FUN=read\_delim, ';', escape\_double = FALSE, col\_names = FALSE, trim\_ws = TRUE)  
data.df = do.call('rbind', datalist)  
colnames(data.df) <- c('CellClass','well-position-CellName','lifespan','xorig','yorig','aggregatelifespan','aggstartx','aggstarty','aggstartz','feature2lifespan','f2startx','f2starty','f2startz','well','name','project','initialstate','endstate','endlife','timepoints','censor')  
  
Note: final timepoints must be censored prior to analysis in R and groups must be defined based on Well or Features. The 'mutate' command is useful for both.  
  
**Example code for survival analysis in R(Studio) after data import:**  
# see this website for tutorial: https://www.datacamp.com/community/tutorials/survival-analysis-R#third  
  
library(readr)  
library(survival)  
library(survminer)  
library(dplyr)  
library(ggplot2)  
  
#Set censor values to binary, and binaraze the event start time to stratify. '9999' is used in ImageJ macro to indicate absence of event.  
dataP.df <- dataP.df %>% mutate(censor\_group = ifelse(censor=='censored', 0, 1))  
  
#Create cutoff for group assignment (in timepoints). Cutoff is 24 in this example.  
cuttoff <- 24  
  
#Assign cells with event before cutoff to the 'Early' group, and cells with events after or no events to the 'Others' Group (Aggregates in this example).  
dataP.df <- dataP.df %>% mutate(EarlyAggregate\_group = ifelse((aggstartz =cuttoff) & (aggstartz !=9999), 'Early', 'Others'))  
  
#Make the grouping variable 'EarlyAggregate\_group' into a factor.  
dataP.df$EarlyAggregate\_group <- factor(dataP.df$EarlyAggregate\_group)  
  
#Censor cells that sruvive to final timepoint creating censor\_group\_Corr, the final censor variable. Max timepoint is 85 in this example.  
maxtimepoint <- 85  
dataP.df <- dataP.df %>% mutate(censor\_group\_Corr = ifelse(lifespan==maxtimepoint, 0, censor\_group))  
  
#Select dataframe for graphing  
data.df <- dataP.df  
  
#Begin survival model, time = survival time, event = censoring  
surv\_object <- Surv(time = data.df$lifespan, event = data.df$censor\_group\_Corr)  
surv\_object  
fit1 <- survfit(surv\_object ~ EarlyAggregate\_group, data = data.df)  
summary(fit1)  
  
#PLOT RESULTS  
ggsurvplot(fit1, data = data.df, pval = TRUE) + ggtitle('Survival Plot')  
  
#Plot of cumulative hazard  
ggsurvplot(fit1, data = data.df, pval = TRUE, fun = 'cumhaz')  
  
#Full plot options  
p <- ggsurvplot(fit1, data = data.df,  
conf.int = TRUE,  
pval = TRUE,  
fun = 'pct',  
risk.table = TRUE,  
#facet.by = data.df$OtherVariables,  
size = 1,  
linetype = 'strata',  
palette = c('#E7B800',  
'#2E9FDF'),  
legend = 'bottom',  
legend.title = '',  
legend.labs = c('Early','Others'))  
  
p <- p + ggtitle('Survival Plot')  
p  
  
# Cox proportional hazards  
fit.coxph <- coxph(surv\_object ~ EarlyAggregate\_group,  
data = data.df)  
ggforest(fit.coxph, data = data.df)