

## 1. PROJECT

-----

Title: Applying Experimental Evolution Techniques to Cancer cell lines.

Dates: 02/16 – 08/18

Funding organisation: University of Reading

This dataset was created as part of a PhD project undertaken at the University of Reading.

## 2. DATASET

-----

Title: Measurements of broad sense heritability of generation time and motility in cancer cell lines.

Description:

Experiments were conducted over 7 months to record the motility and generation time of four cancer cell lines – HeLa, HT1080, MCF7 and MDA-MB-231 - in cell culture. Cell lines were cultured in vitro on a well plate for 24 hours before being placed into the environmental chamber of a Nikon TiE Timelapse Microscope. Multiple points within every well were chosen and videoed over a period of 72 hours. These videos were then analysed and cell motility and division tracked using MtrackJ and ImageJ. Records were kept of which cells divided and the track number assigned to each daughter cell. This allowed comparison between and within cell families so the heritability of the cell traits, motility and generation time, could be estimated. Broad sense heritability was estimated as the slope parameter of an ordinary least squares regression between related cells. Three cell-cell relationships were tested, mother-daughter, sister-sister and cousin-cousin. Our results showed that cell motility has high and significant heritability for all cell lines tested whereas generation time showed little or no significant heritability for the cell lines tested. This is consistent with expectations that traits closely linked to fitness will in general have lower heritability. These results are important for cancer evolution as heritability (a prerequisite for natural selection to be able to act) has never before been measured for any traits in cancer cells.

Publication Year: 2018

Related Publication: [In preparation] Wass, A., Butler, G., Taylor, T. B., Dash, P. R. & Johnson, L. J. Cancer cell lines show high heritability for motility but not generation time. (2018)

Creator(s): Ana Wass & George Butler

Organisation(s): University of Reading

Rights-holder(s): Ana Wass & George Butler

### 3. TERMS OF USE

-----

Copyright Ana Wass & George Butler 2018. This dataset is licensed by the rights-holder(s) under a Creative Commons Attribution 4.0 International Licence:  
<https://creativecommons.org/licenses/by/4.0/>.

### 4. CONTENTS

-----

#### File listing

##### Dataset1\_Cell\_Tracking\_Data.csv

This dataset was originally created in Microsoft Excel 2013 but has been saved as an .csv file for general use. This dataset contains information about the motility and generation time of cell lines gathered through analysis of live cell imaging *in vitro* over a 72 hour period.

##### Dataset 2\_Data\_selection.rdata

This dataset was generated using R and is coding that sorts through the .csv file. It sorts through the .csv file and determines the number of offspring at each generation. It then selects families matching the pre-set criteria and cuts the family size so only the first 4 generations are included.

##### Dataset3\_Heritability\_analysis\_generation\_time.rdata

This dataset was also generated using R. This coding estimates the broad sense heritability of generation time by calculating the slope parameter of an ordinary least squares regression between related cells.

##### Dataset4\_Heritability\_analysis\_speed.rdata

This dataset was also generated using R. This coding estimates the broad sense heritability of cell speed by calculating the slope parameter of an ordinary least squares regression between related cells.

### 5. METHOD and PROCESSING

-----

Cell lines were obtained from ATCC. Cells were cultured in their lab adapted media; MEM with 10% FBS, 1mM Sodium Pyruvate, 2mM glutamine and for HeLa, MCF7 and HT1080 cells, 1% NEAA. Cells were cultured onto a 24-well plate 24 hours before live cell imaging started. Six wells per cell line were distributed around the plate. Five points were chosen at random within each well and images taken at 20 minute intervals over a period of 72 hours using a Nikon Tie Timelapse Microscope. NIS software was used to convert the images into a video file for each point. This was done on three separate occasions. Cells within the videos were then tracked using using ImageJ and MtrackJ<sup>12</sup> software. This tracking data has been edited and saved as the .csv file. The Cell line, cell density, repeat, video, fate, family generation and parent columns have all been added in manually. The other columns were automatically generated from ImageJ and MtrackJ as part of the tracking data.

Dataset1\_Cell\_Tracking\_Data.csv contains the column headers listed below.

Column Title	Description
cell_line	The cell line the data came from.
cell_density	The density at which the cells were seeded at the start of the experiment. This is consistent across all experiments at 5000cells/well.
repeat	The experiment the data was generated from. This experiment was repeated 3 times.
video	Relates to the raw data. Each point chosen within a well was named numerically. When the NIS software created the timelapse videos, these videos were given a file name identical to the points they filmed. This allowed the researchers to go back and identify the video files the cell tracking data originates from. These video numbers are unique within a repeat but not between.
TID	The ID number a cell track is given. Within a video these numbers are unique. However the first cell in each video is always labelled '1' and TID numbering continues sequentially. Each new TID represents one cell.
PID	The number assigned to each point within the TID track. When a cell is being tracked, the centre of the cells position is marked, the next frame in the video is then loaded and the centre of the cell marked again. Each of these points makes up the whole cell track and each is given a unique ID number – the PID. The PID within a cell track is unique, but the numbering resets for a new cell.
x	The calibrated x co-ordinates of the PID. The unit of length in this data set is in microns.
y	The calibrated y co-ordinates of the PID. The unit of length in this data set is in microns.
time	Recorded in seconds. It is the time of the experiment when the PID was created.
length	Measured in microns. This is the length of the cell track from the first point of the track to the current point (inclusive). The value of this quantity for the last point in a track equals the total length of the track.
d2s	The distance from the first point in the track to the current point (inclusive), it is measured in microns and is always less than or equal to the length of the track
d2p	The distance from the current to the previous point of the track. Also measured in microns.
well	Identifies which well in the 24-well plate each video was taken from.
Fate_txt	Whilst tracking the cells a record was kept of what happened to each cell – its 'fate'. This column shows the text format of that result. There were four possible 'fates' – blank: the cell did not divide before the experiment finished. d: the cell divided. died: the cell died. mos: the cell moved off screen and could no longer be tracked.

Fate_Nr	This shows the same information for a cell as Fate_txt but in a numerical format. There were four possible 'fates' – 0: the cell did not divide before the experiment finished. 1: the cell divided. 2: the cell died. 3: the cell moved off screen and could no longer be tracked.
generation	All cells were given a generation number. Cells seen at the start of the video are generation 1. If a cell divides its daughter cells belong to the next generation, generation 2. This continued with each cell division so new daughter cells belonged to the sequential generation from their mother.
family	This was created to give an ID number common to each cell family. All clonal descendants deriving from the same progenitor cell have the same family ID number.
parent	Shows which cells are descendants of each other. If a cell divided a record was kept of its TID as well as that each daughter cell was assigned. This allowed us to input the TID of the parent cell allowing identification; if 0 then this was the first cell seen on screen.

The .csv file contains data on the cells tracked in all three experiments and has not been filtered or statistically tested.

All statistical analysis was performed in R<sup>3</sup>. The generation time for a given cell is the time taken for a full cell cycle from cytokinesis to cytokinesis. The curvilinear speed of a cell is the total Euclidean distance of the path travelled by the cell, divided by its generation time in hours. All cells used in our statistical analysis were observed for a complete cell cycle. Cells that were present at the beginning of the video, those that moved off screen, died or had not divided by the end of the tracking period were therefore excluded. These exclusions led to cell families of different sizes. Statistical testing showed no significant differences in cell speed or generation time between families containing different numbers of cells or the different wells within an experiment. The number of generations used in the statistical analysis was also capped to 4. This was to allow the dataset to run on a less powerful machine as well as the very small n number above this generation.

The broad sense heritability was then estimated as the slope parameter of an ordinary least squares regression between related cells. Three different cell-cell relationships were possible, mother - daughter cell (progenitor cell and its first generation clonal descendant), sister – sister cell (first generation clonal descendants of the same progenitor cell) and cousin – cousin cell (second generation clonal descendants of the same progenitor cell). Where a cell has multiple daughters or cousins their mean value was used. This was done separately for both cell speed and generation time.

#### References:

1. Schindelin, J., et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676-682 (2012).
2. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671-675 (2012).
3. R Core Team. R: A Language and Environment for Statistical Computing. <https://www.R-project.org> (R Foundation for Statistical Computing Vienna, Austria 2018).

