Senior Project proposal:  
new statistical methods to predict patterns of regeneration in flatworms

# Background

A [group of Michael Levin](http://ase.tufts.edu/biology/labs/levin/) (Tufts University) recently discovered that during development and regeneration, many types of somatic cells (not only neurons) change their resting membrane potentials, and communicate to each other using spatial gradients of these membrane potentials (Adams Levin 2013). This is a very new, interesting, and totally unstudied finding, as nobody really thought about resting membrane potentials as something that cells could use to *communicate* to each other; they were always perceived as a relatively boring housekeeping property of a cell. Levin's group demonstrated that by manipulating the resting potential they can change development of an organism; for example, make tadpoles develop a third eye on their back part: some crazy frankensteinian stuff right there! (Blackiston Levin 2013).

Looking for a simpler model to study this communication, Mike Levin decided to use planarians (flat worms). The cool thing about flat worms is that they can regenerate large parts of their body (including their heads). What Levin found is that if you cut the head of a planarian off, and then disturb the slow electrical communication between epithelial cells, planarians frequently grow a "wrong" type of a head (shaped similar to a head from a different, but closely related species), in what seems to be a stochastic manner (Emmons-Bell 2015). Levin also developed a way to image membrane potentials of somatic cells in a worm using voltage-sensitive dyes.

So now

1. we know that the pattern of these potentials is important;
2. we can image them;
3. we have a way to disturb these patterns,
4. and we know the ultimate outcome after disturbance (the type of a head the worm ends up regenerating).

It's time to figure out the connection between the distribution of potentials during regeneration, and the outcome!

See (Durant et al. 2016; Lobo Levin 2015) for a longer restatement of this problem, and (Lobo Beane Levin 2012) for a review.

# The goal of this project in one sentence

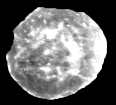
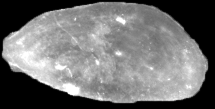
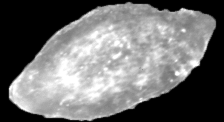
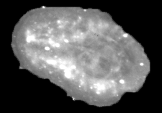
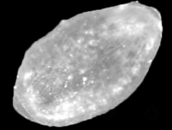
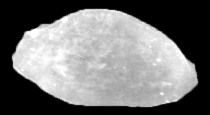
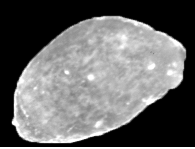
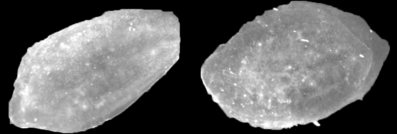
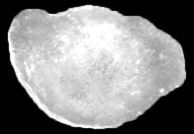
To find a mathematical way to predict which head a worm would regenerate, given a spatial distribution of potentials during regeneration.

# Dataset

We have access to a set of images (voltage-sensitive die photographs, TIFF format) from worms that ultimately end up regenerating different heads. For each head type (4 different types) we will have 10-30 (more if needed?) photos from worms that *normally* grow a head of this type (because it is a normal head type for this species). We also have about the same number of photos from worms that were disturbed during regeneration, and grew a head of this type "by mistake" (even though it is not a normal head shape for them). In summary:

* 8 groups (4 normal, 4 of worms that were manipulated and stochastically ended up with a head of a certain shape)
* 10-30 images per group (closer to 10 for now, but will grow)
* high-res image of each worm (see below)

## Examples of images we are talking about:

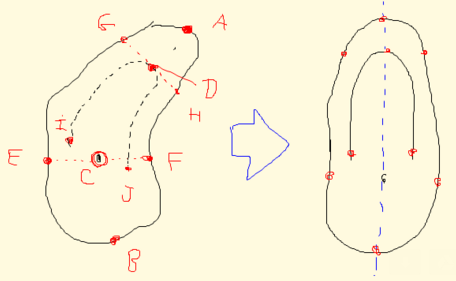
         

# Proposed approach

## 1. Map the worm

While planarians are cool because they are simple and regenerate easily, they are also bendy, squishy, and don't want to stay still. Therefore all photos above have different shapes, sizes and orientations. The first step would be to quantify the distribution of potentials (glowing in images above) along the worm body invariant to the position of this worm.

Instead of adjusting the mapping procedure to the worm shape, I propose to adjust the worm shape to the mapping procedure by transforming it into an "ideal worm", similar to how it is done for faces, maps, or fMRI images (see [Talairach coordinates](https://en.wikipedia.org/wiki/Talairach_coordinates)):



Mathematically these two approaches are more or less equivalent, but there are two benefits it transforming all worms to the same spatial template, and then measuring them, instead of "bending" and adjusting the measurement approach for each image. The first benefit is that it will make it easier to run statistical approaches on the whole image (in relatively preserved resolution) without having to come up with ad-hoc rough regions of interest. The second benefit is that by generating a "mugshot" for each worm, we'll be able to compare these worms visually, which humans are normally good at. With "transformed mugshots" we may be able to see some patterns that are hidden when we look at the original untransformed images. We will also be able to visualize the results of any subsequent statistical analysis better.

In other words, by complicating step 1 a bit, we make step 2 much easier, and much more fun, as we'll be able to run all statistics on a pixel level, between similar images of same resolution each.

From the practical point of view, a student would have to write a program with a graphical user interface, that would allow the user to:

1. Read and display a TIFF file
2. Provide a GUI to mark the nose and the tail of the worm, as well as the midline and some chords crossing these midline. The midline should be a Bezier curve, and chords should (probably) be orthogonal to the midline.
3. Once the worm is marked up, the program should transform the image into an "ideal" rectangular map of given resolution (similar to the original resolution, but unified for all images) using vector transform with bicubic resampling. It should then save the image.
4. We should also be able to save and load mark-ups, in case we need to adjust them later

## 2. Analyze images

Once we have a library of transformed images, we would apply to them some statistical approaches, trying to guess which parts of the pattern are most stable, which are most variable, and which are most informative. Possible approaches:

* Build an average worm for each type, and a map of most variable areas
* Run PCA (or any other type of factor analysis, such as ICA or non-linear multidimensional scaling), and look at the best ways to describe the natural variability of images. Which elements of the pattern come together? Make a bunch of images to illustrate the answer to this question.
* Once the dimensionality is reduced, cluster the worms; analyze variability between the worms.
* Correlate images with head outcomes?

**Optional**: In principle, we could also try to improve the mapping by finding the best adjustment between any given transformed image and a "target image" (an average of several best images?). If the "adjustment transformation" is smooth and restricted in both complexity and "amplitude", it could improve subsequent statistical analysis without overfitting the model. For example, one could consider a class of "parabolic" transformations fixed at both ends [xnew = x + x\*(width-x)\*a] and look for the value of a that minimizes the average squared distances between the transformed image and the target image. Or something to this effect.

## 3. Build and verify a predictor

Can we predict the head type using the pattern observed during regeneration? Possible approaches:

* Primitive (low-hanging-fruit run-of-the-mill approach, we should definitely start with it): PCA (or similar technique) -> clustering -> map head types onto clusters -> see whether the distribution is random (through reshuffling -> empirical distribution -> p-value)
* Bayesian approach (quantify probabilities of getting each head type given a certain pattern; reverse the problem, find "most typical", or perhaps even exaggerated pattern for each ultimate head type)
* Information theory approach?
* Machine learning approaches?

In terms of training and testing the predictor we could:

* Train / analyze "normal" worms and see whether these predictions work for "disturbed" worms
* Train / analyze one set of "disturbed" worms, and then verify the model on a testing set

Regardless of the chosen approach, we should also make sure to get a comparison between "normal worms" and "disturbed worms"-based models.

Ultimately, we would visualize our findings similar to how it was done in my recent stats paper (Ciarleglio Khakhalin et al. 2015).

# Bibliography

Adams, D. S., & Levin, M. (2013). Endogenous voltage gradients as mediators of cell-cell communication: strategies for investigating bioelectrical signals during pattern formation. *Cell and tissue research*, *352*(1), 95-122.

Blackiston, D. J., & Levin, M. (2013). Ectopic eyes outside the head in Xenopus tadpoles provide sensory data for light-mediated learning. *The Journal of experimental biology*, *216*(6), 1031-1040.

Ciarleglio, C. M., Khakhalin, A. S., Wang, A. F., Constantino, A. C., Yip, S. P., & Aizenman, C. D. (2016). Correction: Multivariate analysis of electrophysiological diversity of Xenopus visual neurons during development and plasticity. *eLife*, *5*.

Durant, F., Lobo, D., Hammelman, J., & Levin, M. (2016). Physiological controls of large‐scale patterning in planarian regeneration: a molecular and computational perspective on growth and form. *Regeneration*, *3*(2), 78-102.

Emmons-Bell, M., Durant, F., Hammelman, J., Bessonov, N., Volpert, V., Morokuma, J., ... & Levin, M. (2015). Gap Junctional Blockade Stochastically Induces Different Species-Specific Head Anatomies in Genetically Wild-Type Girardia dorotocephala Flatworms. *International journal of molecular sciences*, *16*(11), 27865-27896.

Lobo, D., Beane, W. S., & Levin, M. (2012). Modeling planarian regeneration: a primer for reverse-engineering the worm. *PLoS Comput Biol*, *8*(4), e1002481.

Lobo, D., & Levin, M. (2015). Inferring regulatory networks from experimental morphological phenotypes: A computational method reverse-engineers planarian regeneration. *PLoS Comput Biol*, *11*(6), e1004295.