A new computational method to quantify 3D image data and to detail changes in morphological structure and spatial relationships during nervous system development

Morgan Schwartz*[1], Jake Schnabl*[2], Benjamin S. Baumer[3], and Michael Barresi[1,2]

[1] *Department of Biological Sciences, Smith College*
[2] *Department of Molecular and Cellular Biology, University of Massachusetts, Amherst*
[3] *Program in Statistical and Data Sciences, Smith College*

January 28, 2018

Abstract

Research in developmental biology has relied on the analysis of morphological phenotypes through qualitative examination of maximum intensity projections that surrender the power of three dimensional data. Statistical methods to analyze visual data are needed, particularly to detect subtle phenotypes. In addition, these methods would best serve the community if they could leverage all the data contained within 3D datasets that are becoming common with the advent of sophisticated microscopy techniques. One barrier to achieving statistical power in image analyses has been the misalignment of spatial relationships between different images. We have created a program for biological image analysis that enables quantification and statistical analysis of 3D multichannel signals that are positioned around a well-defined structure. Our method enables description of the biological structure using a mathematical model that aligns and compares different samples while also accounting for individual variation. We demonstrated the utility of this program by quantifying the phenotypes associated with post optic commissure (POC) and spinal cord development following manipulation of axon guidance cues and astroglial ablation respectively. Our method has successfully quantified a severe non-midline crossing phenotype in the you-too (Gli2-DR) mutant as well as revealed more subtle POC phenotypes in the astray (robo2) and twitch-twice (robo3) mutants that have gone unreported in previous studies. We have also adapted our method to quantify variation in zebrafish spinal cord structures following the ablation of astroglial cells, suggesting this program can be widely applicable. We are currently building this method into a user-friendly, open source program that the community can use to similarly quantify 3D, multichannel datasets, which will provide statistical rigor and novel insight often lost in the qualitative inspection of subtle phenotypic changes.

1 Introduction

The field of microscopy has advanced rapidly since its inception. Recent advances in computer science, computer aided design, and physics have worked synergistically with biology and microscopy to facilitate rapid development of new biological microscopy techniques and instruments over the past two decades. These advances have given rise to a new generation of computer-driven high-resolution microscopes and computer-aided image processing tools like ImageJ and Amira, which

*Contributed equally.

have allowed researchers to investigate fine and complex biological structures in unprecedented detail. Despite the rise of computer-based aids, biologists still primarily rely on innate human pattern recognition to analyze microscope images, but our abilities are rapidly becoming inadequate as data quality and biological phenotypes increases in size and complexity.

Recognizing these challenges, microscopists have sought to leverage recent computational advances in order to introduce greater quantitative rigor and analysis to their work (Xia et al., 2013). While quantitative metrics have slowly made inroads into the field of biological microscopy their implementation has been hampered by the complex, subtle, and highly-variable structures that the field studies and a lack of generally accepted analytical metrics. There have been attempts to respond to this challenge by implementing qualitative classifications that can serve as a basis for statistical analysis, but these approaches have been limited by the ability of the human eye to discern difference and hindered by the natural human tendency towards bias (Barresi et al., 2005). Moreover, attempts at quantifying these descriptive measures fail to provide sufficient statistical power in the face of the subtle phenotypes and morphologies that would most benefit from increased imaging resolution and clarity. These challenges are further magnified by the field's tendency to distill complex images down to 2 dimensions, usually by generating a maximum intensity projection (MIP), which facilitates presentation of the data, but sacrifices much of the complexity and relational data contained in the image. The reduction of three dimensions to two results in a loss of spatial information that is key to detecting subtle differences in biological structures.

However, there have been significant strides made in the last decade in the field of quantitative image analysis, reviewed in Xia et al., 2013, which highlights the potential that new analytical techniques afford the field. Currently, there remains no unified method of analysis, or program of first resort, though ImageJ may be a strong contender for the title, which enables image quantification. This state of affairs is intuitive, as an image may contain many types of information, how bright is the signal, where is the signal, which signal is real, how many sources of signal are there, to name a few, and it is left to the investigator to discern which data is of interest. This has led to a proliferation of analytical methods and programs which attempt to address specific facets of image based questions. This has led to the development of programs which can track cells through time (Amat et al., 2014), techniques to describe cell morphologies (Kriegel et al., 2017), programs which can identify neurons from confocal data sets and build 3D models (Kamali et al., 2009), and even programs which can quickly identify and track live swimming zebrafish based on fish morphology (Chen et al., 2011). Despite these critical advances, new programs and techniques to examine 3D structures still compress complex data into 2 dimensions (Kriegel et al., 2017), lack the ability to utilize biological replicates, and do not enable statistical comparisons between experimental and control samples.

We describe a novel methodology and present a new program which preserves 3 dimensional (3D) information, retaining the spatial relationships of structural components of collected image data, and enables quantitative and statistical comparison of normal and mutant 3D structures that retain a consistent gross morphology. Our methods utilize image data collected from multiple biological replicates to identify regions which are statistically different from control images of the same structure to isolate regions of gross or subtle difference whose features are otherwise not readily identifiable utilizing qualitative analysis. This methodology addresses 4 main challenges in image analysis that we have identified as primary limitations in quantitative image analysis. The first of these are the noisiness of biological data, which may exist due to sample preparation or limitations introduced by the imaging system used to collect the data. The second problem lies in variation introduced due to sample preparation or real biological differences which complicate comparisons between samples. Third, 3D structures are challenging to examine or present, leading many researchers to utilize MIPs, but the loss of structural information and relationships makes investigation of structures or features associated with structures more challenging. Fourth, the complexity and subtlety of biological structures has often lead to the use of qualitative analysis, even if statistics are applied post-hoc, which lack the statistical rigor and power of directly applied quantitative metrics. To overcome these challenges we instituted an image processing protocol and developed *Cranium* to normalize variation in signal intensity due to bleaching or labeling quality, and align, normalize, and quantify changes in spatial relationships between structural components

of the POC. Here we present our new method of 3D image analysis that is capable of quantitative statistical comparisons of complex structural and spatial relationships both in the POC and other biological structures such as the spinal cord.

We demonstrate the use and utility of this program, *Cranium*, in analysis of the zebrafish post optic commissure (POC). The POC is a bundle of midline crossing diencephalic axons, projected towards the midline from ventral rostral clusters of neurons (Figure 1). POC axons form a tightly clustered (fasiculated) bundle as they cross the midline in association with a "glial bridge" composed of Gfap labeled glial cells which precede POC axon crossing (Figure 1)A-C. Visual inspection shows that Gfap is closely associated with the POC though the nature of this relationship is greatly obscured in MIPS; moreover, the sporadic and highly variable nature of Gfap labeling makes it difficult to evaluate changes in Gfap localization. Quantification or analysis of the structure of the POC has proven challenging due to the variable nature of each image collection due to image artifacts, labeling artifacts, and normal biological variation. In traditionally presented MIPs of coronal sections of the POC, the variable structure of the POC and variable Gfap labeling compound data loss in MIPs so that positional information relating the glial bridge to the POC, like whether glial labeling is deep in the collection or right on top of the POC, is lost. The complex structure of the POC and its components makes quantifying changes resulting from mutagenesis and gene knock outs challenging, as changes in glial position, or axon density are hard to discern against the backdrop of natural variability.

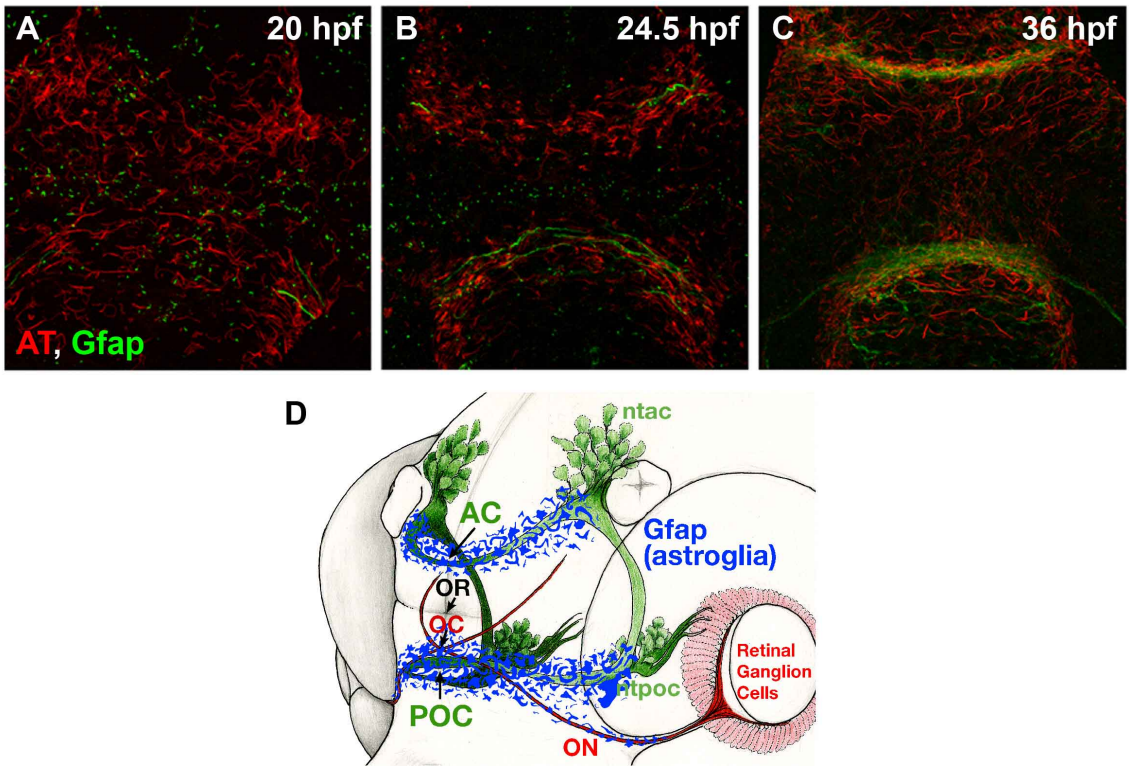

Figure 1: The post optic commissure (POC) is formed by midline crossing axons in concert with a structure of glial cells called the glial bridge.

2 Design and Implementation

We have developed a novel methodology to addresses to 4 major limitations of image analysis that we have identified. These major limitations can be summarized as image noise, biological and sample based misalignment, loss of sample dimensionality, and lack of quantifiability. We have utilized

the open source program, Ilastik, which employs a training based machine learning, to address the first of these challenges, image noise. Our program then preforms principal component analysis to align commissures between samples, reducing misalignment artifacts. We have implemented a novel 3D coordinate system which preserves image dimensionality normally lost in MIPs. Our program then reduces the points identified by the program as belonging to the structure to a set of landmark points that describe the shape and distribution of signal corresponding to the structure. Finally, using our landmark system, we are able to identify and quantify structural differences and changes in signal distribution between wild type and mutant commissures.

2.1 ilastik creates a probability based binary dataset

Our approach to 3D image analysis is centered around the presence of a core biological structure onto which the analysis will be anchored. In the case of the POC, the core structure is made up of POC axons labeled by anti-acetylated tubulin (AT; the structural channel). Any secondary labels, in this case, anti-Gfap, will be processed relative to the core structure.

To address the challenge of noise in image analysis, we are utilizing the interactive machine learning program, ilastik, to distinguish signal from noise (Sommer et al., 2011). ilastik generates a probability value ranging from 0 to 1 for each pixel indicating the probability that it is true signal (Figure 2). A 0.5 probability cutoff is applied to select a set of points that represent true signal for each channel. This probability based method enables confident selection of points relying on statistical significance as opposed to intensity thresholds, which can end up excluding real signal in fainter images. In particular, our implementation of ilastik helps overcome the challenge of differential intensity, which may result from bleaching or poor labeling. Additionally, ilastik can be trained to remove spurious labeling from contaminants and secondary sources of signal through the z stack of an image collection. The output of this method is an image which contains only points of interest, and significantly reduces the number of points which need to be processed in later methods . Importantly for our purposes of axon quantification, ilastik processing reduces signal contamination in the anti-acetylated tubulin channel caused by labeling of cell primary cilia while preserving signal from individual axons. Additionally, it strengthens and clarifies Gfap signal by eliminating the variability in intensity.

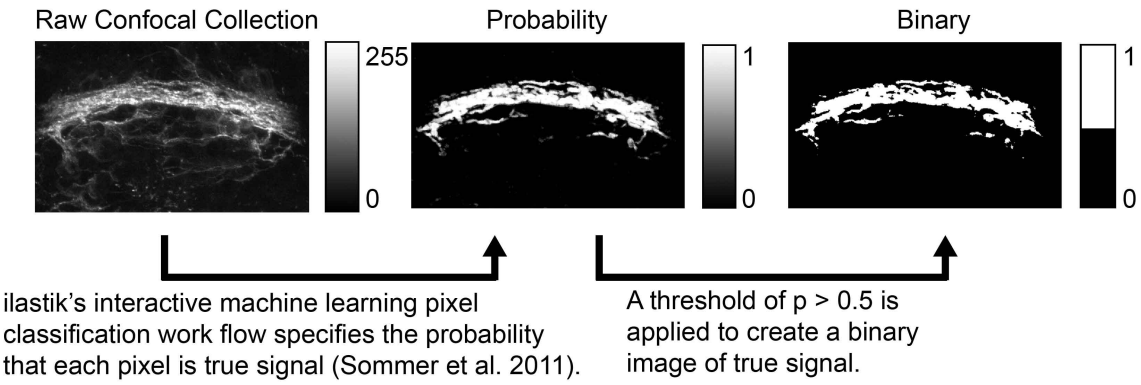

Figure 2: ilastik's machine learning pixel classification work flow enables the selection of a binary set of points that represent true signal after eliminating noise and variable intensity.

2.2 Principle component analysis aligns samples on biological axes

Biological samples, including the POC, frequently vary in how they are collected on the microscope leading to individual differences in how the sample is oriented in 3D space. This artifact of collection means that samples can not be directly overlaid in 3D space without prior alignment. To overcome this challenge, we have implemented principle component analysis (PCA) to robustly identify a

consistent set of biologically meaningful axes in aniosotropic 3D biological samples, in which each dimension is proportionally different in size. This approach is automatic, but easily subjected to user inspection to catch and revise errors.

Our approach relies on a core structural channel, which is used to calculate the PCA transformation matrix. Any secondary channels are transformed according to the matrix calculated for the structural channels. After processing images in ilastik, we have generated a binary dataset of points in 3D space that represent true signal in the channel of interest. In order to prevent fine structures from interfering with the gross morphology of the sample, we apply a median filter twice to smooth the structure and remove fine processes. PCA identifies the axes in the dataset that capture the most variability (Figure 3). Since biological structures frequently maintain consistent proportions, PCA can identify a set of three axes that are biologically meaningful and consistent between samples (Figure 4). In order to ensure that all samples are in the same position following the alignment process, we fit a polynomial model to the data and identify a centerpoint, which is translated to the origin.

In the case of the POC, the relative X axis which spans the medial-lateral dimension of the embryo is consistently identified as the first principle component because it spans a longer range of values than the other dimensions (Figure 4). Additionally, the dorsal-ventral axis of the commissure is collected in the relative Z axis and typically has a greater range of values than the Anterior-Posterior axis captured in the Y axis. A parabola ($z = ax^2 + bx + c$) is used to describe the shape of the commissure and the vertex of the parabola is used to identify the center of the data to position at the origin. Following PCA alignment, the commissure lies entirely in the XZ plane with the midline of the commissure positioned at the origin. After PCA and alignment is completed on the structural channel, AT, the same transformation and translation is applied to any secondary channels.

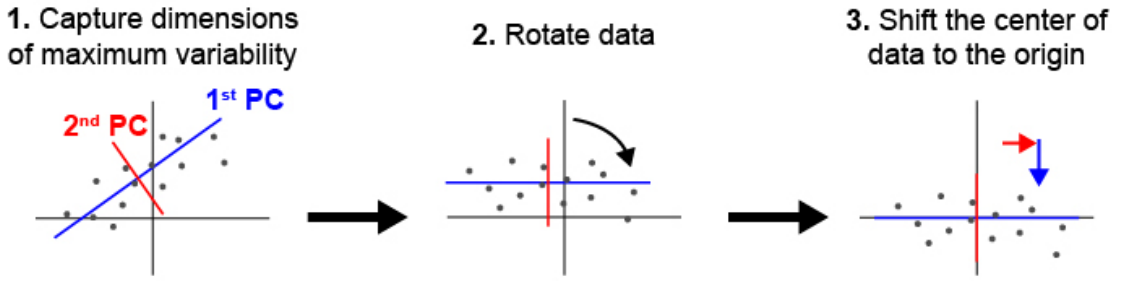

Figure 3: Principle component analysis is applied to n-dimensional data to identify axes that capture the most variability in the data. 1) In this two dimensional example, the first principle component (PC) is identified and the second PD is oriented perpendicular to the first. 2) The data is rotated so that the first PC is horizontal and the second PC is vertical. 3) Finally, the data is shifted so that the center of the data is at the origin.

2.3 Cylindrical coordinates defines the position of signal relative to the biological structure

Microscopic techniques necessarily collect data in the three primary spatial axes, x, y, and z, and as technology has advanced, the resolution of each of these axes has increased. While these advances have resulted in increased precision and accuracy in signal localization in the three spatial axis, they have not resulted in increased understanding of the spatial relationship between signal from different channels. This limitation is largely due to the common methodology of creating two-dimensional maximum intensity projections from three-dimensional image stacks. MIPs frequently give the impression that signal in two channels may be spatially related when they are actually

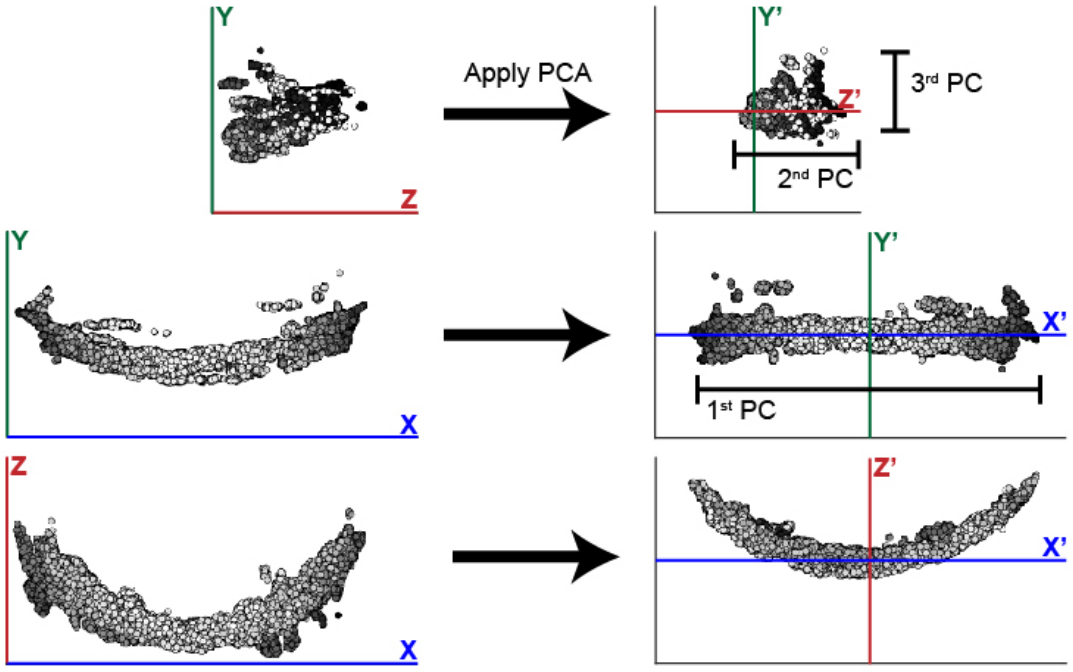

Figure 4: After applying PCA to a single POC sample, the arc of the commissure lies in the new XZ plane with the midline positioned at the origin. In each 2D projection, the intensity of each point corresponds to depth in the third dimension.

separated by in the missing third dimension. The previously described qualitative method of analyzing biological images have left researchers to make ambiguous judgment calls about this critical metric of spatial relationships.

In order to define a coordinate system that contains inherent biological meaning, we have implemented a cylindrical coordinate system that defines the position of each point of signal relative to the biological structure as a whole. First, we return to the polynomial model that was used to describe the biological structure during sample alignment. With regards to the model, each point is defined by its distance from the model (R), its angular position around the model (θ) and its distance from the midline along the model (α) **??**. These new coordinates serve as parameters that contain information about the shape and composition of the structure. From this point, the data is able to be qualitatively and quantitatively analyzed, using 1 and 2 dimensional methods.

The cylindrical coordinate system that describes the POC is built off of a parabola that describes the shape of the commissure in the XZ plane: $z = ax^2 + bx + c$. α describes the position of points relative to the midline or periphery of the commissure. θ captures the dorsal-ventral or anterior-posterior position of the point. Finally, R describes how far a point is from the commissure, which captures the degree of fasiculation of the commissure.

2.4 Landmarks serve as a representative set of points for comparison between samples

The conversion to cylindrical coordinates enable each point to encode biological information; however, the large number of points in each sample makes it difficult to draw a direct comparison between samples. In order to reduce the data to a form that facilitates direct comparison, we have drawn inspiration from the landmark analysis that is commonly used in morphological studies (). When used in morphological analysis, landmarks are assigned to distinct points on the shape of interest and the change in position between samples is used to quantify different shapes. Typically landmarks are assigned by an expert with domain knowledge of the structure in question; however,

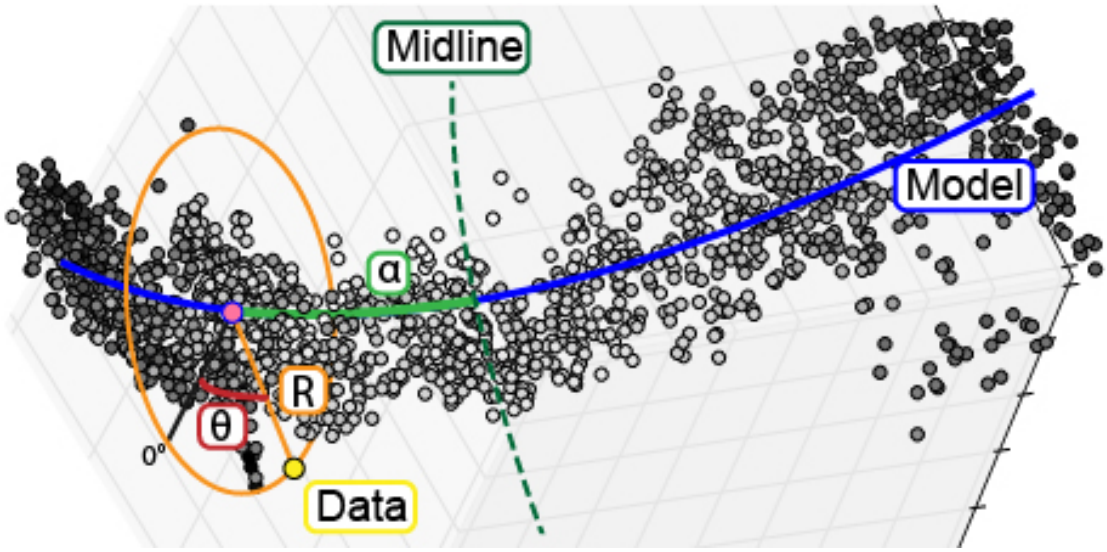

Figure 5: To enable analysis of data point relative to a biological structure, points are transformed from a cartesian coordinate system (x,y,z) into a cylindrical coordinate system (α,θ,R) defined relative to the structure.

we have chosen to calculate an automatic set of landmarks distributed across the structure in order to avoid introducing bias due to expectations about where biological differences should emerge.

In order to calculate a set of landmarks that describe the set of points that form the structure, we divide the data into a set of bins in the α and θ axes. Within each bin, we calculate the 50^{th} percentile of the points in the R dimension. The final position of each landmark is determined by the midpoint of the α bin, the midpoint of the θ bin and the 50^{th} percentile of R. The θ dimension is divided into 8 bins, each of $\frac{\pi}{4}$ radians. This division enables resolution along the anterior-posterior and dorsal-ventral biological axes without oversampling. The number of bins along the α dimension is determined by optimization.

Figure 6

In order to identify the size of an α bin that ensures a representative sampling of the data, we calculated two measures of variance. First, the variance is calculate for each bin and its adjacent neighbors. This type of variance between bins decrease as the number of α bins increases. In order to counter this trend, we calculated the variance between samples for each bin, which decreases as the number of α bins decreases. By selecting the number of α bins that minimizes both types of variance, we identify the appropriate number of bins for the sample. In the case of the POC, 30 bins minimizes variance between bins and between samples.

2.5 Random forest classification identifies novel regions of phenotypic difference

The previous set of sample processing steps have created a framework in which biological structures can be directly compared to others in order to identify differences. However, in the case of phenotype analysis, a system is needed to identify which landmarks represent the greatest difference between samples. We have turned to machine learning methods to train a classifier that distinguishes sample groups based on a set of landmarks as predictors. In order to insure that the final model retained information about the importance of landmark predictors, we selected a random forest classifier.

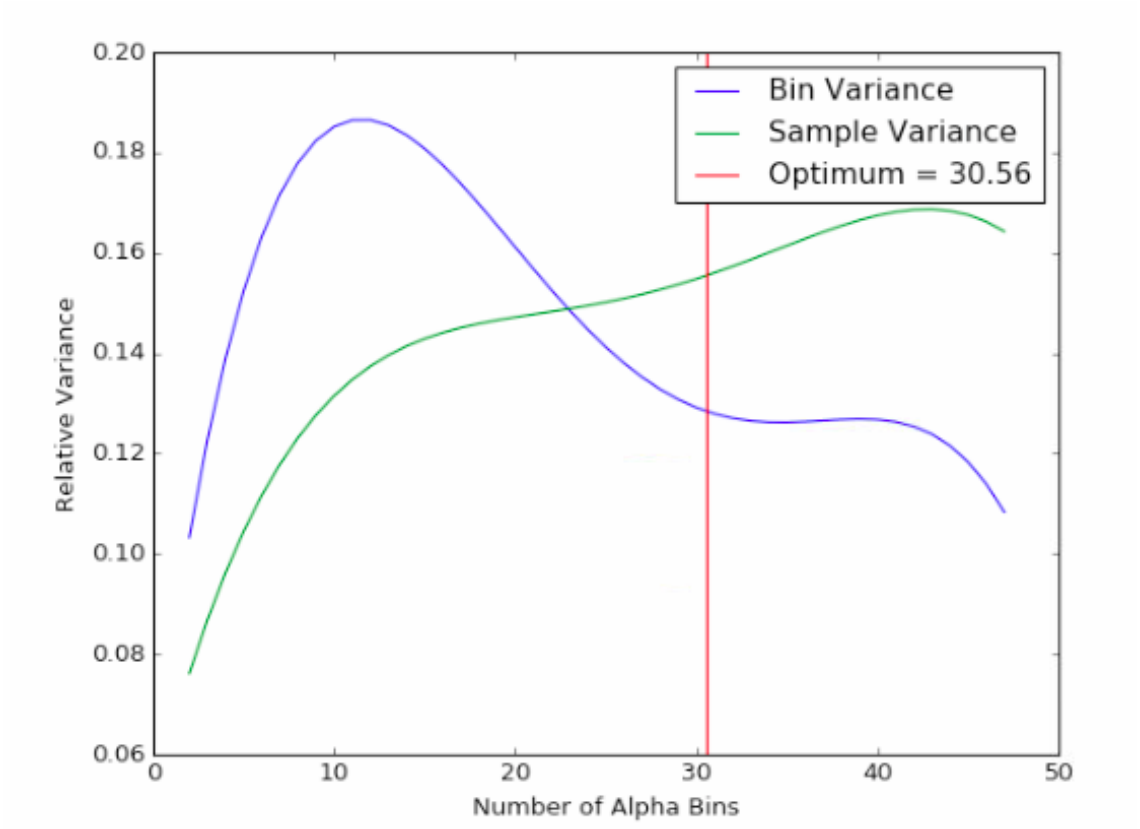

Figure 7

Calculating landmarks for the POC resulted in a set of several hundred landmark points that greatly exceeded the number of samples available for training the model. In order to reduce the number of landmark predictors to less than the number of samples, we applied principle component analysis and performed a parameter sweep to identify the minimum number of components that captures 100% of the variability present in the landmark set. Following transformation of the data according to PCA, the random forest was trained with a set of 500 trees sampling from $\sqrt{p}$ predictors using the out-of-bag score to assess error. The resulting random forest classifier includes a score for each component that captures how heavily that component was weighted in training the model. Subsequently, each component can be visualized as the set of all landmark points with each point receiving a coefficient that determines its contribution to the component. By analyzing the landmark contribution to components that heavily influence the random forest classifier, we can identify novel regions of biological difference.

3 Results

3.1 Testing The Model: POC formation in the axon guidance mutant You-Too

Initial proof of principal testing of the quantification program was performed on acetylted tubulin labeling of the POC, which forms a clear and continuous structure in the zebrafish forebrain. Cranium was tested using wild type and you-too mutant POCs, because the dramatic visual reduction in midline crossing and commissure formation present in the you-too mutant **??** has already been previously described and therefore provided a concrete comparison for our method. Previous methods, as described in (Barresi et al 2005), utilized a phenotypic binning metric, defining

6 levels of commissure development, from wild type fully fasciculated fully crossed commissures to non crossing commissures. This metric was a great improvement over previous methods, but was still reliant on observer skill, robust phenotype, and utilized discrete steps which are difficult to accurately assign. The Cranium program provides an unbiased and continuous analysis of the POC, and represents a next step forward in structural analysis of microscopic structures. Our first objective in development and testing of Cranium was determining its ability to detect significant disturbances in the amount of signal present at the midline in you-too mutants as compared to wild type embryos. This problem poses a significant challenge to more classical methods of image quantification, as the you-too mutant as exhibits an increase in cilliary labeling in the AT channel, which methods of whole image pixel intensity calculation register as real signal. Moreover, you-too mutants appear to have more non-commisural axon labeling, axons which extend deeper into the forebrain, are non fasciculated, and only part way extend towards the midline, which in MIPs are counted as pre-commisural axons.

3.1.1 Cranium set up for you-too analysis

Our method to eliminate background labeling employed the open source program, Ilastik. Ilastik uses user input and a machine learning algorithm to assign a probability to each pixel that the signal observed at that point is signal of interest. Ilastik utilizes several image processing filters to assist in detection of "real" signal, of which we used all image filters from .5 pixels to 5 pixels, and then initiated user training on a minimum 5 wild type and 5 you-too samples, with additional samples used until the program successfully recognized commissural axons at all z depths while excluding cilliary labeling. The output of this training was a probability based image, which we evaluated by comparing MIPs from multiple samples from before and after training **??**. After training, we observed that the MIPs from wild type and you-too mutants trained data exhibited less background labeling, though there was also a decrease in visibility of some lone axons **??**. We determined that fasciculation and axon density, being hallmarks of commissure formation, were not adversely impacted by the training, and as such, decreased visibility of non-commissural axons would not adversely impact quantification of commissure formation and extent.

Using the probability based images, we applied a binary filter set at $p < .25$, so that only highly likely "real" signal would be examined. This probability was chosen because .25 and .5 were identical from a structural perspective, but .25 presented significant gains in image analysis speed. In testing, values as low as $p < .1$ were not drastically different from .25 or .5, and also presented even more significant gains in speed, but did result in a loss of representation by smaller and thinner axonal processes, and this was not used in final analysis **??**. From this point, the cranium program, including PCA, binning, and analysis were conducted as previously described in 2.

3.1.2 You-too axon analysis

Utilizing our landmark analysis method, we observe that there is significant decrease in the number of positive signal pixels at the midline, which measures the amount of commissure formation, in you-too embryos as compared to wild type. This decrease in pixel number occurs in all bins, (Figure 8), suggesting that this decrease is not due to mis-localization of the midline crossing point. Interestingly, cranium also shows a significant decrease in signal across the entire length of the tract of the POC, not just at the midline, suggesting that you-too embryos do notform a commissure, a fasciculated bundle of axons, but that they either do not project axons towards the midline at all, do not exhibit fasciculation en route to the midline, or some middle ground between these extremes. The presence of points in the tract of the POC but not at the midline seen in (Figure 8) suggests that the fewer axons overall are projected toward the midline compared to wild type, and that very few of those axons make it to the midline.

To examine the structure of the commissure and address whether you-too commissural axons are de-fascuculated en route to the midline, we also looked at the distance (r) of signal from the prospective location of the tract of the POC. Interestingly, we show that commissural axons in the tract of the POC are not significantly more defasciculated in you-too as compared to wild type

(Figure 9). However we do see that axons at the prospective location of the POC midline are de-fasciculated, and given that axons are not significantly de-fasciculated in the tract of the POC (Figure 9) it is likely that the observed de-fasciculation is due to displacement of commissural axons from the tract of the POC.

This methodology provides increased resolution of significant differences between wild type and you-too embryos. Previously published methodologies, as published in Barresi et al. 2005, were able to describe the failure of commissure formation in you-too mutants, but were unable to discern the significant decrease in commissural axon crossing. Moreover, you-too mutants were characterized by their defasciculation, however we show that you-too commissural axons travel along a tract that is not significantly wider than that of wild type embryos. These results demonstrate increased accuracy and precision over previously published methodologies.

3.2 Subtle Phenotypes: Gfap Distribution in the Tract of the POC

The ability of the program to substantiate and describe in greater and more precise detail the known commissure formation phenotype of the you-too mutant led us to attempt to utilize cranium to identify more subtle forebrain phenotypes which have eluded analysis. Barresi et al. 2005 noted a potential distrubance of Gfap distribution around the POC in you-too embryos as compared to wild type. It has been suggested that the you-too mutant phenotype results in disrupted glial bridge formation, which subsequently impacts commissure formation but the dispersed labeling of Gfap has made reliable quantification of this phenotype challenging. Gfap, which is a well described marker for glial cells, which are the predominant neural stem cells of the zebrafish forebrain, is a common marker in brain and spinal cord analysis. Most often it is used to label zones of radial glial proliferation or localize glial positions relative to other cell types like endothelial cells, in the context of the blood brain barrier. Gfap is a structual intermediate filament in glial cells however, and its labeling is only suggestive of where any individual glial cell is. Moreover, Gfap labeling is diffuse, chaotic, and 3 dimensional complicating more rigorous analysis. Indeed, in the zebrafish forebrain, Gfap labeling is found throughout the forebrain, including deep below the POC, which greatly complicates analysis of Gfap distribution when utilizing MIPs.

In order to test whether cranium is able to quantitatively describe differences in Gfap labeling in you-too embryos, we utilized the AT data from wild type and you-too embryos to generate a location for the tract of the POC. We then calculated the amount of Gfap signal and the distance (r) of Gfap signal around the commissure. We observe that there is a significant increase in amount of Gfap signal in the tract of the POC which is most pronounced in (Figure 8) F,G. Moreover, we show that there is very little Gfap signal present at the midline in comparison to the periphery in either wild type or you-too mutants. Interestingly we see that Gfap signal is significantly more closely associated with the tract of the POC in you-too mutants in all axes(Figure 9), despite there being more signal in the tract in you-too mutants as seen in (Figure 8) F,G.

(Figure 9) (Figure 8).

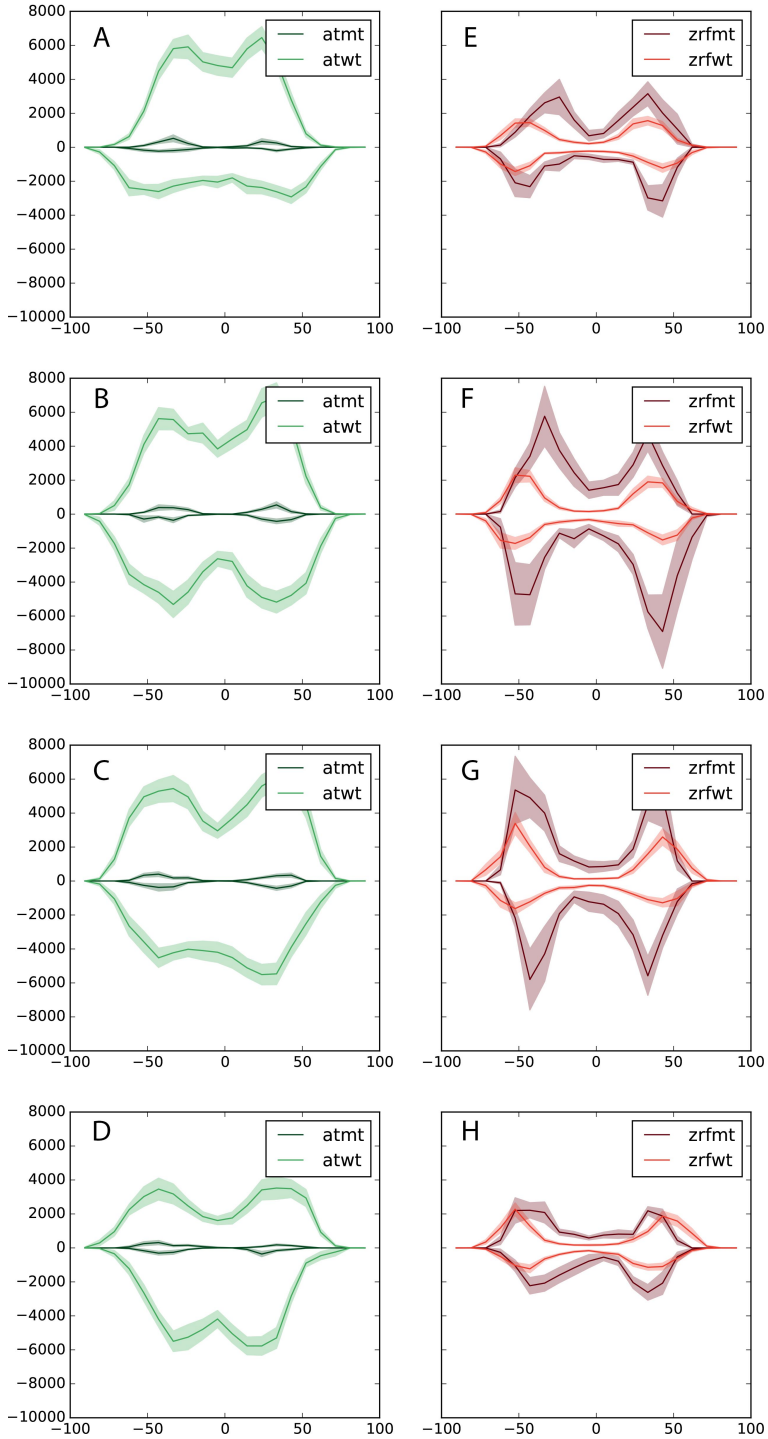

Figure 8: A-D Acetylated tubulin (AT) labeling is significantly reduced in Yot embryos as compared to WT including at the midline. E-F Gfap signal in Yot mutants is significantly increased at the periphery and the tract of the POC, but is similar to WT at the midline.

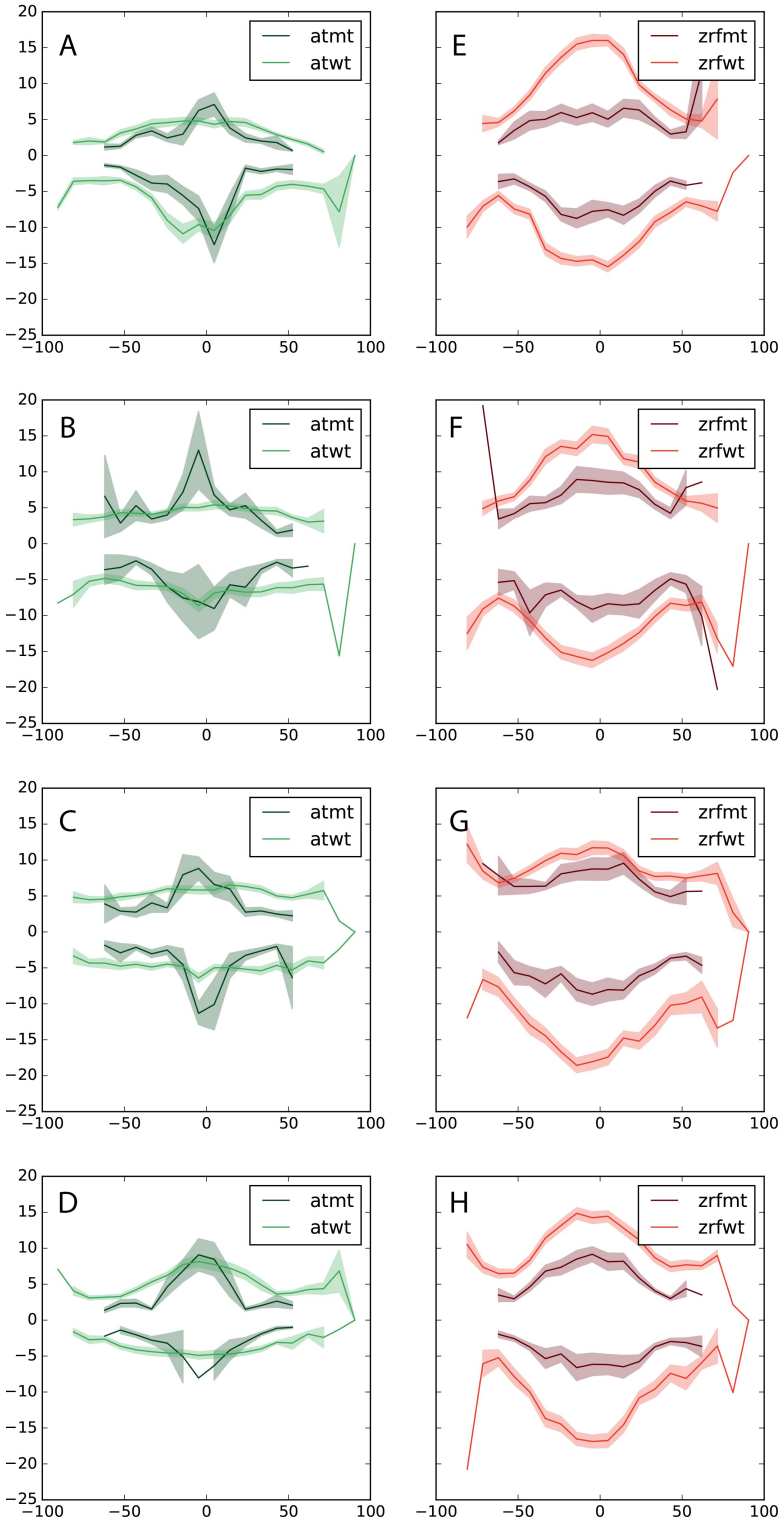

Figure 9: A-C Acetylated tubulin (AT) labeling is significantly dispersed compared to WT at the midline. D AT labeling in Yot mutants shows no significant perturbations. E-F Gfap signal in Yot mutants is more tightly restricted to the prospective area of the tract of the POC

4 Discussion

5 Materials and Methods

5.1 Zebrafish Husbandry

Fish lines were maintained in the Smith College Animal Quarters according to Smith College Institutional Animal Care and Use Committee (IACUC) and AAALAC regulations. Groups of 12-15 fish were housed in 1L tanks on a fish holding system with recycling water at a standard temperature of 28.5 - 30.0°C and a 12 h light-dark cycle. Adult zebrafish were maintained on a diet of dry fish food (Aquaneering, San Diego, CA) and brine shrimp (Artemia International, Fairview, TX). Baby zebrafish were fed a liquefied mixture of hatchfry and spirulina until they were of reproductive maturity (approximately 3 months). For data collection, embryos were collected in the morning after a male and female fish had been separated into a divided mating chamber the prior evening. The morning of collection, dividers were pulled and fish normally mated within the first 30 minutes, and healthy fish lay multiple times within the next few hours. Embryos were harvested and placed in embryo media (EM) comprised of 5mM NaCl, 0.17mM KCl, 0.33mM CaCl2, 0.33mM MgSO4 and 0.00003% methylene blue. Embryos were staged by morphology and age (Kimmel et. al, 1995).

5.2 Wild-type and mutant fish lines

Wild type embryos were from the AB lines provided by C. Lawrence (Harvard University) and Tübingen from ZIRC (Eugene, OR). You-Too lines were generously provided by the Karlstrom Lab (Amherst, MA).

5.3 Immunocytochemistry

Fish were set and left to lay embryos over night and were collected no later than noon the next day. Embryo age was assessed between 4 and 6 hpf the day of collection and before 24 hpf the following day, to ensure accurate staging of embryos using the ZIRC zebrafish embryonic staging chart. Embryos are raised in E3 with .2% methylene blue, and .003% Phenylthiourea (PTU) is added 1:10 in E3 at 24 hpf for embryos fixed later than 30 hpf. Embryos fixed for POC immunocytochemistry were fixed at 27.5-28 hpf. Embryos fixed for spinal cord immunocytochemistry were fixed at 36 hpf. Embryo fixation was performed with 4% formaldehyde diluted in .025M phosphate buffer (PB), and fixed for 2 hrs at room temp, or 4 degrees C for 16 hours for both 27hpf and 36 hpf embryos. Embryos are then washed out of fix, first with 3 rinses of .1M PB, followed by 3 5 minute washes in .1M PB. Fixed embryos are then rinsed once with 50% Methanol, and then immersed in 100% Methanol for no longer than 20 minutes. Fixed embryos are then immersed in 100% acetone, 4 minutes for 27.5-28 hpf embryos, 7 minutes for 36 hpf embryos. Embryos are then rinsed 2 times with 100% methanol, and then washed for 5 minutes with 50% methanol. This is followed by 3 rinses and 3 5 minute washed in phosphate buffered saline with 2% v/v triton x-100 (PBS-Tx). 36 hpf embryos are also treated with .5% collagenase for 30 minutes, and a subsequent 3 rinses and 3 5 minute washes in (PBS-Tx). Fixed embryos are then blocked for 1 hr at room temp in PBS-Tx with 2% w/v bovine serum albumin fraction V, 1% v/v dimethylsulfoxide, and 10% v/v normal goat serum (block). POC immunocytochemistry was performed using mouse Igg2B anti-Acetylated tubulin, and mouse IgG1 anti-ZRF-1 bade up in block, while spinal cord immunocytochemistry was performed using rat anti-dsred, rabbit anti-gfp, and mouse IgG1 anti-ZRF-1 made up in block. This primary step was conducted at room temp for 2 hours, or 4 degrees C for 16 hours for POC imaging, or only 4 degrees C for 16 hours for spinal cord imaging. Fixed embryos were then rinsed 3 times and then washed 3 times for 15 minutes each with PBS-Tx before being placed back in block for 30 minutes. Embryos were then placed in secondary, POC samples in anti-mouse IgG1 488 and anti-mouse IgG2B 647 conjugated antibodies made up in block, and spinal cord embryos in anti-mouse IgG1 647, anti-rat 594, and anti-rabbit 647 conjugated antibodies made up in block. POC secondary was conducted at room temp for 2 hours, or 4 degrees C for 16 hours for POC

imaging, or only 4 degrees C for 16 hours for spinal cord imaging. Samples were then rinsed 3 times and washed 3 times for 15 minutes in PBS-Tx, and then placed and allowed to fully sink in 70% glycerol 30% PBS.

5.4 Antibodies

The following primary antibodies were used: mouse acetylated tubulin (AT, Sigma, 1:800), mouse zebrafish radial filament-1 (Zrf1, Sigma, 1:4), rabbit glial fibrillary acidic protein (GFAP, Sigma, 1:400), mouse monoclonal DsRed (Clontech, 1:500) and rabbit polyclonal DsRed (Clontech, 1:500). Secondary antibodies included: goat anti-rabbit 488 (Invitrogen, 1:200), goat anti-mouse 647 (Invitrogen, 1:200), goat anti-mouse IgG-594 (Sigma, 1:200), goat anti-mouse IgG2b Alexa 647 (Invitrogen, 1:200).

5.5 Confocal Microscopy

Embryos used for POC imaging were decapitated and mounted in 70% glycerol with the POC oriented closest to the glass cover slip. Embyros were then imaged on a Leica SP5 scanning confocal microscope at 63X with a 1.5 optical zoom. Images were collected at 1024 by 1024 pixels with a line average of 4. A z stack was collected of the POC of each embryo with a step size of 0.21 um. Laser intensities were chosen for long captures, with the argon laser power set to 25% with an intensity of 12%, the 594 laser was set to 80%, and the 633 laser was set to 20%. The stacks were captured at 600 hz, with bidirectional scanning active, and configured to a phase difference of -34.75. A minimum of 20 um of POC was captured from each embryo, up to a maximum of 35 um, in order to capture the full extent of the tracts of the POC. Stack capture times were around 15 min, with captures around 20 min showing bleaching of the samples.

5.6 HDF5 Files

Following imaging on the confocal microscope, images are saved in '.lif' files. Each sample is opened in Fiji using the Bioformats plugin, cropped to eliminate background in X and Y, and rotated to position anterior as up. Each channel is isolated and saved as an individual HDF5 (.h5) using the HDF5 plugin for Fiji.

5.7 ilastik

Confocal images of the POC, consisting of Gfap and AT channels were imported into ilastik and filters relevant to their channels were selected as appropriate. Positive signal and background were trained, with strong emphasis placed on judicious application of the background channel to remove spurious and diffuse labeling. Probability images were exported and compared to the originals to ensure fidelity of signal and success in noise removal.

5.8 Preanalysis data processing

Following processing in ilastik, 'Probabilities.h5' files are read as a 4D array ([zyxc]). ilastik saves the data with two channels: signal and background. In order to distinguish between these channels in the data array, we assume that the channel that contains more points with a probability of greather than 0.1 than the number of points less than 0.1 is the true signal channel. After reading the data as an array and selecting the appropriate channel. Each point is saved to a dataframe as a row with x, y, and z values obtained from the point's position in the dataframe. At this time, each point's xyz position is scaled to account for the size of the voxel collected by the microscope. Finally, a threshold is applied to the dataframe containing all points to select only points with a probability of less than 0.5. This set of points will serve as the set of points that represents the structure of interest.

5.9 Sample Alignment

In order to ensure that sample alignment is not negatively impacted by any potential fine structures or noise, we perform two preprocessing steps on the structural channel that are use exclusively for alignment. First a new dataframe is created as described above, but a threshold of 0.25 is used to select a more stringent set of points. Second, a median filter using a radius of 20 is applied to the thresholded data twice in order to smooth out noise on the surface of the structure. For wildtype samples, principle component analysis (PCA) is applied to the processed data using all three dimensions. Following transformation of the structural channel, the first principle component is assigned to the x axis, the second to z and the third to y. The transformation matrix that is calculated for the structural channel is also applied to any secondary channels. *You-too* samples undergo the same processing steps, but PCA is performed on only the y and z axes and the first and second components are reassigned to the z and y axes respectively. The severe phenotype present in *yot* samples can interfere with the alignment process if the x axis is included.

5.10 Sample Centering

Following PCA samples are aligned in a consistent orientation, but their positions in 3D space are still variable. In order to center all samples at the origin, a polynomial model is used to represent the underlying shape of the data and to identify a consistent center. After alignment, the POC lies in the XZ plane and forms a parabolic structure. We fit a quadratic equation, $z = ax^2 + bx + c$ to the structural channel by minimizing the squared error. The vertex of the parabola is calculated and the dataset is shifted in order to position the the vertex of the origin. The y position of the data is determined by calculating the average y value and shifting it to the origin. The necessary translation is calculated based on the structural channel and then applied to any secondary channels.

5.11 Cylindrical Coordinates

In order to convert the xyz coordinates of each point into a cylindrical coordinate system defined by α, R, and θ, each point $P(x_p, y_p, z_p)$ is considered in the XZ plane and a secondary point $P_m(x_m, 0, z_m)$ is identified on the previously calculated model that minimizes the eucidean distance between the point and the model. The pythagorean theorum is used to calculate the distance between P and P_m, $r_p = \sqrt{(z_p - z_m)^2 + (y_p - y_m)^2}$. α is defined by calculating the distance along the model between P_m and the vertex $V(x_v, 0, z_v)$, $\alpha_p = \int_{x_v}^{x_m} \sqrt{1 + (2ax + b)^2}$. Finally, θ is calculated as follows: $\theta_p = \arctan \frac{y_p - y_m}{z_p - z_m}$.

5.12 PSI Files

After transforming the data to cylindrical coordinates, the dataset is saved to a '.psi' file according to PSI Format 1.0. Each point is assigned an ID number and the following values are saved: x,y,z,α,R,θ.

5.13 Batch Processing

In order to minimize processing time, five samples are processed in parallel through the workflow described above.

5.14 Landmark Calculation

In order to enable direct comparison of samples, each sample is reduced to a set of representative landmark points. The data is divided into eight bins around θ, each spanning 45°, and into a bins along alpha. For each landmark, the number of points in this subset is calculated and the 50^{th} percentile of the R values of the data is identified. The number of bins along α is determined by minimizing both the variance between samples and the variance between adjacent bins.

Each landmark, $B_{i,j,s}$ is defined by the the index of the α bin, the index of the θ bin and the sample number.

The variance between adjacent bins is calculated as follows for each possible number of α bins (a):

$$Bvar(a) = \frac{\sum_{s=1}^{n}(\frac{\sum_{i=1}^{t}(\sum_{j=1}^{a}(\frac{\sum_{k=j-1}^{j+1}(B_{i,k,s}-\bar{B})^2}{2})))}{t \times a})}{n}$$

The variance between samples is calculated as follows for each possible number of α bins (a):

$$Svar(a) = \frac{\sum_{i=1}^{t}(\sum_{j=1}^{a}(\frac{\sum_{s=1}^{n}(B_{i,j,s}-\bar{B})^2}{n-1}))}{t \times a}$$

The optimal value for a is identifying the value that minimzes the sum of $Bvar(a)$ and $Svar(a)$. For wildtype samples of the POC, the optimal number of α bins is 30.

5.15 Testing the difference between wildtype and *yot*

5.16 Random Forest Classification

In order to identify significant differences between sample types, we are using a classifier to identify the most important landmarks. The previously calculated landmarks are used as predictors for the classifier. However the number of landmark predictors (p) exceeds the number of samples that are used to train the model (n). In order to return to $p < n$, PCA is applied to the set of landmarks and a parameter sweep from 1 to 100 is performed to identify the number of components that captures 100% of the variability in the data. Following transformation to the optimal number of components, a random forest classifier is trained using a maximum number of $\sqrt{p}$ features per tree and 500 trees in the forest. The out-of-bag score is used to measure the error in the classifier.

After the classifier is trained, the importance score of each component is obtained and used to identify components, which had the most influence in distinguishing wildtype from *yot* samples. Furthermore, each component can be visualized as a heatmap of the landmark set with the color representing the weight of each landmark for the component. This visualzation can be used to identify landmarks that are notably different between wildtype and *yot*.

References

Amat, Fernando et al. (2014). "Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data". In: *Nature methods*.

Barresi, Michael JF et al. (2005). "Hedgehog regulated Slit expression determines commissure and glial cell position in the zebrafish forebrain". In: *Development* 132.16, pp. 3643–3656.

Chen, Shiye et al. (2011). "Automated analysis of zebrafish images for phenotypic changes in drug discovery". In: *Journal of neuroscience methods* 200.2, pp. 229–236.

Kamali, M et al. (2009). "Automated identification of neurons in 3D confocal datasets from zebrafish brainstem". In: *Journal of microscopy* 233.1, pp. 114–131.

Kriegel, Fabian L et al. (2017). "Cell shape characterization and classification with discrete Fourier transforms and self-organizing maps". In: *Cytometry Part A*.

Sommer, Christoph et al. (2011). "Ilastik: Interactive learning and segmentation toolkit". In: *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, pp. 230–233.

Xia, Shunren et al. (2013). "Computational techniques in zebrafish image processing and analysis". In: *Journal of neuroscience methods* 213.1, pp. 6–13.