FISHATLAS: ASSESSMENT OF ORGANOTROPISM DETERMINATION THROUGH IMAGING INFORMATICS OF XENOGRAFTED ZEBRAFISH

APPROVED BY SUPERVISORY COMMITTEE

Gaudenz Danuser, Ph.D. Advisor

James Amatruda, M.D. Ph.D., Co-Advisor

Reto Fiolka, Ph.D., Chair

Rolf Brekken, Ph.D.

Angelique Whitehurst, Ph.D.

Ralph Mason, Ph.D.

# DEDICATION

I would first like to thank my wife, Elizabeth, for her support and listening to me amid a pandemic and many changes over the past three years. I would not be where I am now without her help. In addition, I would like to thank my family for their support during my time in graduate school. Everyone's support during the pandemic, research ups and downs, and a host of other things that life crops up were some of the main reasons I can write this dissertation. Thank you all for your sacrifices in guiding me and allowing me the freedom to pursue something that I love and fascinates me since I was very young: science.

To my grandfather, thank you for sharing your time and stories with me as a nuclear scientist. In many ways, those stories sustained me through difficult times and inspired me to use science to help people. I appreciated you teaching me to be curious, build things, and ultimately what it looks like to dedicate your life to helping people and using science.

To my siblings, thanks for letting me make random calls to you when I was frustrated when something hadn't worked for the 1000$^{th}$ time. While most might have discounted their brother as crazy, you all listened and talked to me.

My great friends Tarek Ghaddar, Logan Scott, and my TAMU Nuclear Engineering Aggies, Park Cities Baptist Church, and so many others thank you for being my friends. When I moved to Texas in 2014, I was unsure what I would face, but I couldn't have expected to have met the people who would impact my life for the better for many years. These years have meant the world to me, and it was because of each of you all.

To my classmates of the incoming class of 2016, Nick Lesnar, Mike Cotten, Alexander LaFrance, and the Danuser + Amatruda laboratories, thank you for helping me be a better person and scientist and listening during failed experiments.

To my scientific mentors, James Amatruda, Gaudenz Danuser: thank you for teaching me and continuing to mentor me as an independent scientist.

I cannot state strongly enough that this dissertation is the product of many people offering their help and support in this project. I would not be the scientist or person I am if it wasn't for collaboration, good people, and trying times to see what you can make or discover.

# FISHATLAS: METASTATIC ORGANOTROPISM DETERMINATION THROUGH IMAGING INFORMATICS OF XENOGRAFTED ZEBRAFISH

By

DAVID HAMILTON SAUCIER

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY
The University of Texas Southwestern Medical Center at Dallas
Dallas, Texas
December 2021

FISHATLAS: METASTATIC ORGANOTROPISM DETERMINATION THROUGH IMAGING INFORMATICS OF XENOGRAFTED ZEBRAFISH

David Hamilton Saucier, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2021

Supervising Professors: James Francis Amatruda, M.D., Ph.D. and Gaudenz Danuser, Ph.D.

Ewing sarcoma patients with metastatic disease have a 5-year survival rate of approximately 28%. The hallmark of this disease is an aberrant transcription factor made by a fusion of Chromosomes 11 and 22 called *EWSFLI1*. *EWSFLI1* expression levels have been correlated with differing responses in cell metastatic propensity, but much remains to be elucidated. Indeed, many current models fail to meet the statistical rigor that is needed for exceedingly spontaneous, rare events like metastasis. To address this need, FishATLAS utilizes zebrafish human cancer cell xenograft images after high fidelity registration using a novel diffeomorphic transformation to display metastatic hot spots of different cancer cell

conditions to begin to grasp the deeper underpinnings of organotropism in vivo with individual cancers. Utilizing a suite of statistical tests, FishATLAS determines at a global whole-fish scale and the local microenvironment, if there are statistically different cell hotspots when comparing two or more different conditions. As it stands, data for EWSFLI1, its target SOX6, a non-transformed cell line NIH3T3, TC32 subclones, and melanoma cell lines have all shown unique distributions of metastatic hot spots. These findings serve as a tool for drug discovery and later environmental re-mapping in FishATLAS by allowing transgenic fish images (such as vasculature and lymphatics) to be overlayed onto any historical data set. These can then be used to determine a given microenvironment's contributions to secondary sites of metastasis. In the case of EWSFLI1 and its target SOX6, there was a marked difference upon shRNA-mediated knockdown that removed a population of hotspots in the upper somitic veins while some were persistent post genetic perturbation. SOX6 shRNA KD data indicates that the somite and inter-somitic arteries are more sensitive for metastatic colonization. Previous studies and our current accumulator data suggest these to be regions with higher oxidative stress, guiding insights for oxidative-mechanistic therapy. These and other conditional accumulations of sparse metastatic hotspots demonstrate the power of FishATLAS as a longitudinal assay of cellular and genetic conditions across all cancers.

TABLE O F CONTENTS

PRIOR PUBLICATIONS

Marchetto, A., Ohmura, S., Orth, M.F., Knott, M.M., Colombo, M.V., Arrigoni, C., Bardinet, V., **Saucier, D**., Wehweck, F.S., Li, J. and Stein, S., 2020. Oncogenic hijacking of a developmental transcription factor evokes vulnerability toward oxidative stress in Ewing sarcoma. *Nature communications*, *11*(1), pp.1-16.

**Saucier, D.H**., 2016. Monte Carlo Irradiation Model For Long-Term Low Dose Environments for Space Life Science Applications (Master's Thesis).

Clanton, R., **Saucier, D.,** Ford, J. and Akabani, G., 2015. Microbial influences on hormesis, oncogenesis, and therapy: a review of the literature. *Environmental research*, *142*, pp.239-256.

LIST OF FIGURES

LIST OF TABLES

# LIST OF DEFINITIONS

EwS – Ewing sarcoma

TME – Tumor microenvironment

DoG – Difference of gaussian

ELS – Ewing-like sarcoma

TDA – Topographical data analysis

DPI- Days post-injection

DPF- Days post-fertilization

PDX- Patient-Derived Xenograft

PNET- Primitive neural ectodermal tumor

LAP- Linear assignment problem

SOX6- Sry box 6 Transcription Factor

EWSR1 – Ewing sarcoma breakpoint region 1

FLI1 – Friend leukemia integration 1 transcription factor

FA- FishATLAS

WT- Wildtype

sh*- shRNA construct

# CHAPTER ONE
# Introduction and Review of the Literature

## *1.1 Ewing Sarcoma*

Ewing sarcoma (EwS) is the second most common bone cancer, with additional occurrence in soft tissue, that has its peak occurrence in adolescence at the age of 11-12, predominantly occurring in people of European descent. As a bone cancer, it is commonly found in the pelvis, femur, tibia, and ribs, and soft tissue in the thoracic wall, gluteal muscle, pleural cavities, and cervical muscles (Gaspar, Hawkins et al. 2015).

EwS is a very aggressive disease. There are about 1.5 cases per million children of European descent annually, 0.8 cases/million for Asian descent, and 0.2 cases/million for African descent. Of these cases, 20% of patients present resistant metastatic disease yielding a very poor prognosis (Gaspar, Hawkins et al. 2015). Upon metastasis, 5-year survival of patients is approximately 30% compared to a survival rate of 70-80% in patients with localized tumors. Unfortunately, even with treated cases, there is a relatively high rate of relapse of 40%, with most of those recurrent patients having drug-resistant disease (Grunewald, Cidre-Aranaz et al. 2018). Current treatments are typically combination therapy of radiation treatment with chemotherapy as well as surgical resection, which are all dependent on the staging and localization of the tumor (Stahl, Ranft et al. 2011).

Ewing sarcoma family of tumors include extraosseous EwS, peripheral neuroectodermal tumors (PNET), and Askin tumors. Although this might indicate a potential neuroectodermal or mesenchymal-based lineage for EwS, nearly 100 years after its discovery, the exact cell of origin remains undetermined (Toomey, Schiffman et al. 2010).

Ewing sarcoma is classified histologically as a small round blue cell tumor. Molecular insight into the origin of the disease began with the discovery of characteristic chromosomal translocations in this disease that generate novel fusions between a FET family gene and an ETS-family transcription factor. The most common fusion is EWSR1-FLI1 which occurs in 85% of EwS cases. A summary of these and their relative frequencies are given below in Figure 1.

| *FET* part | *ETS* part | Fusion gene | Chromosomal translocation | Frequency |
|---|---|---|---|---|
| *FUS* | *FEV* | *FUS–FEV* | t(2;16)(q35;p11) | <1% |
| | *ERG* | *FUS–ERG* | t(16;21)(p11;q22) | <1% |
| *EWSR1* | *FLI1* | *EWSR1–FLI1* | t(11;22)(q24;q12) | ≈85% |
| | *ERG* | *EWSR1–ERG* | t(21;22)(q22;q12) | ≈10% |
| | *ETV1* | *EWSR1–ETV1* | t(7;22)(p22;q12) | <1% |
| | *ETV4* | *EWSR1–ETV4* | t(17;22)(q21;q12) | <1% |
| | *FEV* | *EWSR1–FEV* | t(2;22)(q33;q12) | <1% |
| | *ETV5 (?)* | *EWSR1–ETV5* | ? | ? |
| *TAF15* | ? | ? | ? | ? |

**Figure 1 Ewing sarcoma comes from many different chromosomal fusions, with EWSFLI being the most common at 85%. (Grunewald, Cidre-Aranaz et al. 2018)**

Ewing sarcoma is representative of many pediatric cancers, as it is driven by a single genetic event (the EWSR1-FLI1 fusion) with few additional somatic mutations, in contrast to

adult-type malignancies. As there are no clinically validated inhibitors of EWSR1-FLI1 function, the lack of recurrent cooperating mutations means there are few opportunities for targeted therapy of Ewing sarcoma. Thus, understanding how transcriptional targets of EWSR1-FLI1 may promote its oncogenic function is critical.

Indeed, future Ewing sarcoma treatment could be greatly improved by making a pipeline to allow testing of a large number of candidate target genes and their *in vivo* microenvironmental manifestations. It should also be noted that although this table encompasses different FET and ETS parts, translocations utilizing non-FET parts are a separate class of oncogene transcription factors called Ewing Like Sarcoma (ELS) which comparatively are very rare (Delattre, Zucman et al. 1992).

The specific gene region names for EWSR1-FLI1 are FET-Ewing sarcoma breakpoint region 1 (EWSR1) and ETS-friend leukemia integration 1 transcription factor (FLI1). FET genes encode RNA-binding transcription and splicing proteins that when combined with an ETS transcription family member, provide a means for cell proliferation, differentiation, cell cycle, and apoptosis dysfunction lending to the pathogenic behavior of EwS (Sorensen, Lessnick et al. 1994). These fusion proteins search and bind at de novo gene enhancer GGAA motifs, which could provide a potential target for therapy (Riggi, Knoechel et al. 2014). Currently, a solution to fully exploiting this finding is non-trivial and remains to be seen.

EwS is more prevalent in certain ethnicities, suggesting that inherited germline factors may affect susceptibility to this disease. This is evidenced by its much higher incidence in those of European descent as well as no evidence for environmental risk factors being causal in patients developing EwS. Evidence of disease clustering in families or relatives does suggest a germline-based risk factor to Ewing sarcoma (Crompton, Stewart et al. 2014). Overall, the genetic mutational burden is low at diagnosis, which is common for pediatric cancers. Some characteristic mutations in the STAG2, TP53, and CDKN2A regions are areas of active research interest as exploitable drug targets or therapies (Grunewald, Cidre-Aranaz et al. 2018).

In essence, EwS is a disease that works by reprogramming the epigenome and altering gene regulatory elements that cause a cacophony of epigenetic dysfunction in patients. This presents a challenge in determining how these transcription factors' sequelae affect the overall poor prognosis of patients with metastatic disease.

### 1.1.1 Ewings-like Sarcomas (ELS)

Ewing-like sarcomas (ELS) share commonality with EwS in their pathologies being caused by aberrant transcription factors through gene fusion. Although rarer than EWSR1-FLI1 positive sarcomas, they contain many different subgroups with unique phenotypes that are worth mention for future studies. ELSs are a heterogeneous collection of sarcomas that contain slight differences from clinical EwS and were initially grouped with EwS until 2010. ELS was split at this time due to their lack of a FET-ETS gene fusion. ELS genomics data

were able to split this rare but heterogenous family into three distinct groups (Grunewald, Cidre-Aranaz et al. 2018) :

BCOR-Sarcomas: BCOR-CCNB3, BCOR-MAML3, ZC3H7B-BCOR

CIC Sarcomas: CIC-DUX4, CIC-FOX04, CIC-NUTM1

NATC2 Sarcomas: EWSR1-NATC2

Patients with ELS are commonly treated in a similar treatment regimen to EwS and many research attempts to elucidate optimized treatments are being undertaken but limited by the overall rarity of specific ELS fusion oncogenes. Despite a different classification, many BCOR fusion-positive tumors have similar clinical outcomes to EwS. In contrast, CIC fusion-positive tumors are typically met with poorer outcomes than EwS or other ELS on the list (Kao, Owosho et al. 2018).

ELS constitutes an important population of fusion-positive sarcomas in addition to EwS. Despite having some similarities to EwS, ELS tumors manifest distinct biological and oncogenic mechanisms. Indeed, there is increasing evidence of each being a unique entity further demonstrating the need for wide-reaching genomic assays. Currently, ELS poses a significant challenge in determining treatment for patients(Renzi, Anderson et al. 2019).

### 1.1.2 Biological Background: Transcription Factors

The central driver of Ewing sarcoma, EWSR1-FLI1, is the result of a chromosomal translocation fusion between chromosomes 11 and 22 notated as t(11:22)(q24:q12) resulting

in an aberrant transcription protein utilizing the ETS transcription factor FLI1. Transcription factors, as well as cell regulators, function by remodeling the chromatin storage units, histones, to modulate proliferation, differentiation and other fundamental developmental processes in normal tissues as well as cancer. However, in transformed cells, these highly controlled processes undergo dysfunction and EWSR1-FLI1 can activate pieces of the genome machinery that are normally silent. These oncogene-mediated changes manifest unique biomarkers, the most recent suggestion being IL6 regulation, which may be a dominant promoter of aberrant behavior in EwS (Kondo 2019).

EWSR1-FLI1 targets a specific binding domain motif consisting of a GGAA nucleotide sequence which makes up part of the natural binding domain of ETS transcription factors, however, the oncogenic fusion uniquely targets tandem GGAA microsatellite sequences (Riggi, Knoechel et al. 2014). By binding these GGAA regions, EWSR1-FLI1 engages specific gene promoters as well as epigenetic regulators that explain the overall gene activation and repression patterns of EWSR1-FLI1. Thus, similar to other oncogenic transcription factors, EWSR1-FLI leverages epigenetic chromatin remodeling via dysregulation to select for conditions lending to cancer's progression and survival (Baylin and Jones 2011).

A tempting, promising therapeutic answer to EwS would target these unique GGAA microsatellite regions preventing the oncogenic hijacking of the epigenome without disrupting the normal ETS sequencing targets. However, this is non-trivial and would require

a precision medicine delivery system to EwS cells to block out these regions. This specific treatment remains unsolved. What can be done to progress is to begin to understand the singular and combinatorial effects of EWSR1-FLI's activated and repressed genes and how this relates to a sudden, rapid change in solid tumor to metastatic disease.

The second most common EwS fusion, EWSR1-ERG, is nearly structurally identical to EWSR1-FLI1 and has nearly identical binding regions on the genome (Sorensen, Lessnick et al. 1994) with EWSR1-ERG patients met with a similar prognosis to those with EWSR1-FLI1 fusions. Although there exists a great degree of similarity between EWSR1-FLI1 and EWSR1-ERG, there is not enough information to determine if treatment would work ideally for other fusion-positive sarcomas like BCOR or CIC fusions. Experimentally testing each of these unique cases is key to future treatment pathways and further stresses the urgent need for a high-throughput system to systematically probe each fusions' specific mechanisms.

Several candidate druggable targets of EWSR1-FLI1 have been identified, including polyADP-ribose (PARP). EWSR1-FLI1 transcribed RNA has a recurrent looping structure (R-loops) that could allow for PARP inhibitors to be effective via inactivation of BRCA1 thereby reducing the DNA damage repair of the cancer cell (Gorthi, Romero et al. 2018). The efficacy of PARP inhibitors is now being investigated for EwS and other sarcomas such as alveolar rhabdomyosarcoma (Camero, Ceccarelli et al. 2019).

EwS is a disease of the epigenome. This is established in copious amounts of data showing the minimal genetic mutations that are in EwS tumors (Sorensen, Lessnick et al. 1994, Toomey, Schiffman et al. 2010, Crompton, Stewart et al. 2014, Grunewald, Cidre-Aranaz et al. 2018) and the epigenetic regulators that are targeted by EWSR1-FLI1. Of particular interest is STAG2, an epigenetic regulator that generates de novo enhancers in EwS that dynamically shift the cell to a more mesenchymal-like cell (Grunewald, Cidre-Aranaz et al. 2018). EWSR1-FLI1 selectively activates and represses different regions through these enhancers. Activation through EWSR1-FLI1 works by passing through the closed-chromatin states at GGAA microsatellites thereby increasing exposure of the previously closed DNA to other transcription factors and chromatin complexes. Recently, it is hypothesized that this effect is proliferated through the genome through the interaction of BAF chromatin remodeling to increase GGAA microsatellite recruitment (Boulay, Sandoval et al. 2017).

### 1.1.3 SOX6: Direct Hijacked Target of EWSR1-FLI1

One such additional target through GGAA sequences is the transcription factor SRY-box transcription factor 6 (SOX6). SOX6 is a member of a transcription factor family defined by the presence of a high-mobility-group (HMG) domain. SOX family proteins are responsible for determining cell fate and differentiation. Aberrant SOX6 transcription levels have been linked with gliomas, lung adenocarcinomas, pancreatic cancers as well as others (Guo, Yang et al. 2013, Jiang, Yuan et al. 2018). This transcription factor is responsible for the proliferation and stem-cell differentiation of chondrocytes in bone tissue as well as other

developmental processes. Specifically, in EwS, SOX6 has been established as highly expressed in Ewing sarcoma cells enacting high proliferation and increased mesenchymal phenotype (Marchetto, Ohmura et al. 2020).

Perhaps most interesting is that SOX6 also has a high degree of epigenetic dependence due to its upstream target, EWSR1-FLI1. Constitutive expression of SOX6 in a cancer cell has been directly linked to the amount of exposed intronic GGAA microsatellites on the SOX6 gene (Marchetto and Grünewald 2020). High frequency of GGAA microsatellite exposure is met with higher levels of SOX6 expression, and correspondingly SOX6 mediated phenotypes of proliferation and mesenchymal behavior (Marchetto, Ohmura et al. 2020). This finding is important because it offers a potential explanation of how the epigenome of EwS mediates tumor heterogeneity of SOX6 expression. It is highly speculated that this phenomenon is not limited to SOX6, but other genes containing enriched regions of GGAA microsatellite repeats (Riggi, Knoechel et al. 2014, Boulay, Sandoval et al. 2017).

Precisely how genetic and epigenetic changes in EwS cells affect the interaction of the cells with the tumor microenvironment (TME) requires more elucidation. Indeed, the interplay environment and epigenome could reveal mechanisms of disease progression and metastasis in this disease.

## *1.2 Metastasis Clinical Presentation*

Pediatric cancer treatments have taken huge advances in the past decades seeing patient mortality decreasing by >50% since 1975 (Smith, Altekruse et al. 2014). Correspondingly, patients presenting with a solid EwS tumor with no metastatic disease on average have a 5-year survival of 82%, increasing from 59% in 1975.

Metastasis in EwS is the single biggest predictor of patient outcome. Patients with metastatic disease have a survival of 25% corresponding to a drop of 50% predicated solely on the presence of metastasis (Siegel, Miller et al. 2021). EwS is not unique necessarily in the sense of a higher staging dropping survival precipitously, but in comparison to other common cancers such as melanoma (27%) and breast cancer (28%), it takes a similar trend to more aggressive cancers.

In addition to the critical issue of metastasis, EwS has astonishingly poor outcomes in patient relapse with overall survival of 13% (Anselmino, Rovera et al. 2019). EwS metastatic disease, although a minority in comparison to most cancer cases, constitutes a very serious mortality factor amongst other cancers, with no true development in treatment for decades due to lack-luster responses to modern treatments (Hesla, Papakonstantinou et al. 2021). However, the future holds promise in many new technologies for targeted drug treatment.

Metastatic disease has been attributed to a host of causes including epigenetic shifts, environmental cues, mesenchymal transitions, and selected genetic mutations. The speculated landscape of what causes metastasis is a complex question of the primary tumor's intrinsic genetic heterogeneity and environmental signals. Metastasis in Ewing sarcoma and melanoma present different challenges for FishATLAS and widespread genomic screens.

Specifically, in Ewing sarcoma, metastasis remains poorly understood. Metastasis has been linked to a germline polymorphism in CD99, an Ewing sarcoma clinical biomarker (Manara, Pasello et al. 2018), stress-induced pathways suggesting environmental triggers, pAKT pathway, and most interestingly RAC1 activation by FGF2 in the bone TME (Krook, Nicholls et al. 2014). RAC1 is a potentially insightful connection to melanoma and its metastatic effects mediated by RAC1 mutations (Kamura, Matsumoto et al. 2010, Baldauf, Orth et al. 2018).

FishATLAS's goal is to deterministically look at metastasis through both the lens of genes as well as TME for EwS. Analysis into environmental cues mediated by growth factor starvation or increased hypoxia increase expression of CXCR4 from environmental stressors are points of interest. Observed upregulation of CXCR4-SDF-1 chemokine axis increases the motility of EwS cells to distant cells by increased production of stromal-derived factor 1 (Krook, Nicholls et al. 2014). Interestingly, this environmental signal is both very dynamic and transient making in vivo experimentation potentially confounding outside of the zebrafish system. CXCR4 is both a dramatic, reversible switch hypothesizing cells quickly

revert a distant, more hospitable microenvironment (Krook, Nicholls et al. 2014). CXCR4 effects in vitro seem to be ameliorated via RAC1 inhibitors mechanistically suggesting Rho-GTPases as a mediator of environmental CXCR4 effects (Krook, Nicholls et al. 2014). This makes a strong connection to current research in metastatic melanoma, where RhoGTPases have been linked with modulated actin activity and invasion of cells (Colón-Bolea, García-Gómez et al. 2020).

The second environmental signal for Ewing sarcoma cells is hypoxia, which canonically activates the HIF1a axis to promote angiogenesis and as a sequela, metastasis (El-Naggar, Veinotte et al. 2015). Hypoxia additionally modulates how transcriptional activity happens to lead to HIF1a induction through EWSR1-FLI1 targeting of Dickkopf-related protein 2 (DKK2) (Hauer, Calzada-Wack et al. 2013). Hypoxia is another metric that could be determined in the fish environment via the use of stains on live fish, a relatively simple experiment to determine its potential contributions to the metastatic phenotype.

The last environmental signal indicated in the literature for Ewing sarcoma metastasis was oxidative stress. Increases in reactive oxygen species/oxidative stress increase expression of STEAP1 metalloreducase, increasing Ewing sarcoma invasiveness (Grunewald, Diebold et al. 2012). STEAP1 has been determined to be a direct target of EWSR1-FLI1 that is heterogeneously activated partially contingent upon environmental state/stress (Grunewald, Diebold et al. 2012). This target in addition to HIF1a would be targets of interest to note the interplay of cell and environmental oxidative stress. \

Genetic EWSR1-FLI1 direct transcription targets implicated in metastasis include CAV1, GPD64, EZH2, NPY, TRIP6, and CD99 (Martins, Ordóñez et al. 2011, Grunewald, Cidre-Aranaz et al. 2018, Manara, Pasello et al. 2018). Each of these candidate genes is promising but are likely a piece in the overall landscape of gene stresses and require more experimentation to determine how individually and synergistically each contribute to the EwS metastatic cascade in addition to any compensatory effects if one is modulated.

Each of the above three environmental stressors and suggested genetic targets plays a role in the overall metastatic phenotype of EwS. Research cites an epithelial to mesenchymal (EMT) transition as the critical mechanism for metastasis in vivo (Machado, López-Guerrero et al. 2012). Indeed, the ectopic level of EWSR1-FLI1 expression is set by both intrinsic heterogeneity and environmental signals with higher expression of EWSR1-FLI1 correlated with a higher frequency of EMT (Franzetti, Laud-Duval et al. 2017). However, there is a partially stochastic element to the amount of EWSR1-FLI1 fusion protein contributing to this EMT transition. It is speculated that cells in the primary tumor with more EWSR1-FLI1 protein are more likely to undergo EMT and display metastatic behavior potentially explaining one hypothesis heterogeneity and metastatic frequency (Franzetti, Laud-Duval et al. 2017).

### 1.2.1 Metastasis in Melanoma

In direct contrast to EwS, melanoma metastasis is highly conditioned upon mutational genotype and burden and has one of the highest mutational burdens of any cancer (Sha, Jin et al. 2020). Figure 3 below shows 25 different cancers and their respective mutational burdens.

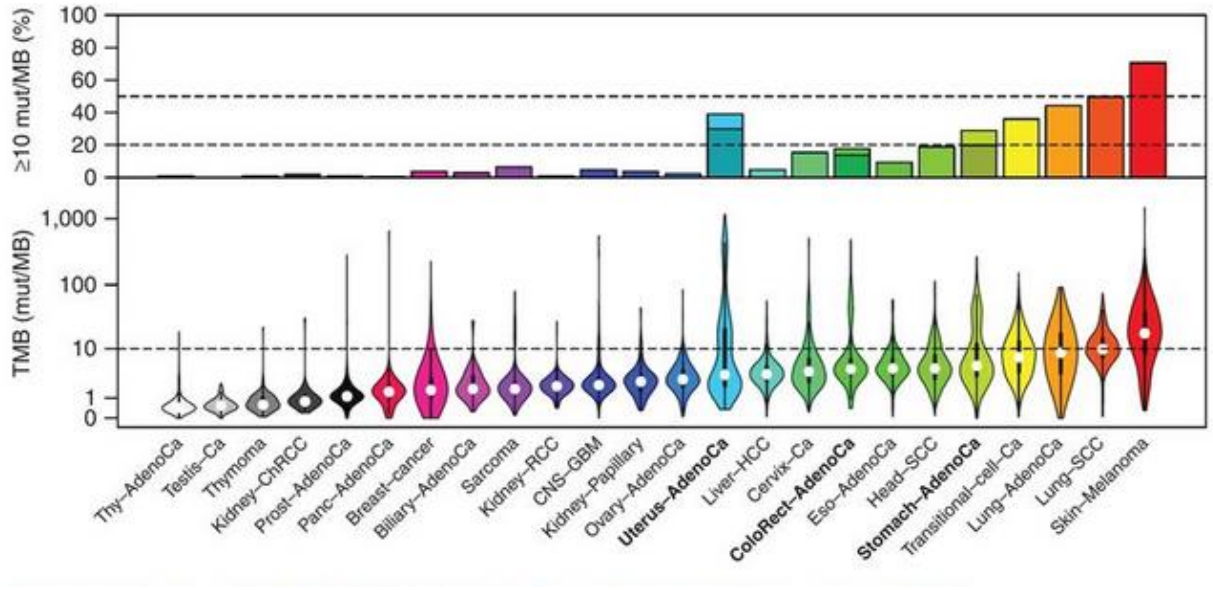



**Figure 2- Melanoma has the highest tumor mutational burden (TMB) than any of the other 25 sampled cancers. (Sha, Jin et al. 2020)**

MV3 and A375 melanoma cell lines have previously been characterized and shown to have different metastatic propensities (Zaritsky, Jamieson et al. 2021)). A375 contains a b-RAF mutation/s and potentially a secondary RacP29S mutation, a powerful regulator of proliferation and actin assembly (Mohan, Dean et al. 2019) that shows higher invasion and metastatic potential. Alternatively, MV3 was hypothesized to gain metastatic propensity due to loss of ECM adhesion, based on in vitro studies in collagen (Friedl, Maaser et al. 1997).

Future experimentation will show which hypotheses are the stronger driver to metastatic behavior. Data provided by FishATLAS provides an opportunity to look at both MV3 and A375 metastatic behavior in vivo.

One of the most prominent hypotheses links metastatic progression in melanoma cell lines to a BRAF V600E mutation, which constitutively activates cellular RAF. It remains controversial if b-RAF mutations directly contribute to metastasis or are a pre-conditioning step that increases the odds of metastasis (Long, Menzies et al. 2011). In clinical data, b-RAF mutational status in melanoma was not correlated with increased metastasis, but positive mutations were associated with lower 5-year survival (Long, Menzies et al. 2011).

In truth, the question of what causes metastasis in melanoma is likely a more complicated answer. Although most literature seems to agree that b-RAF does elicit a change, whether it is a pre-condition for metastasis or if RAF activation is consistently needed has yet to be founded (Libra, Malaponte et al. 2005). In addition, many of the older collagen experiments lack biological context and require experimentation in vivo to fully encapsulate the environmental pressure put on melanoma cells. However, these are well within the capabilities of the FishATLAS pipeline and presented in chapters 3 and 4

## *1.3 Zebrafish Animal Models*

Zebrafish (*Danio rerio*) is a well-established research model organism that offers advantages immediately to computational pipelines. Zebrafish as a model organism in FishATLAS offer immediate benefits and limitations as shown in the following Table 1.

**Table 1 Benefits and limitations of using human xenografts in zebrafish larvae. (Konantz, Balci et al. 2012)**

| Advantages | Limitations |
|---|---|
| • Large numbers of offspring<br>• Transplantation at embryonic stages possible<br>• Permeability of zebrafish to small molecules delivered by the water enables drug screens<br>• External development and existence of transparent lines allow visualization and bioluminescence readouts<br>• No immune rejection in early transplantation settings<br>• Small numbers of cells per animal required for xenotransplantation<br>• Visualization (transparency, transgenic lines, measurable bioluminescence)<br>• Fast readout (hours to days) | • Little knowledge about niche structures and microenvironmental cues<br>• Different biological environment in developing zebrafish for transplanted adult cells<br>• Differences in size (small zebrafish organs/vessels—large human cells)<br>• Different maintenance temperatures (overcome at 35 °C)<br>• Absent organs (e.g., breast, lung) (may be overcome using analogous structures, e.g., gills)<br>• Fewer possibilities for orthotopic transplantation<br>• Limited numbers of zebrafish antibodies available<br>• No adult immuno-permissive zebrafish lines available yet |

Zebrafish is an ideal system to show how the complex pressures of genetics and the TME mesh together to enact metastasis in EwS. There are fundamental limitations with other organisms that are limited by imaging opacity, time-latency, and statistical power through sample size. This type of assay would be both costly and present difficulty in terms of imaging and high throughput in mice.

Zebrafish, in addition, provide ease of transgenic studies that can be used to determine tissue-specific hypotheses using the established GAL4-UAS tissue-gene system that is prevalent in Drosophila melanogaster. Such resources and limited negative tradeoffs are why the zebrafish was chosen as the model organism for FishATLAS.

## *1.4 Zebrafish Xenotransplantation + Genetic Models*

Precedents for zebrafish xenografts were shown starting in 2005 when melanoma cells were xenotransplanted into blastula stage embryos and later observed to determine how they would behave or survive as the embryo developed (Lee, Seftor et al. 2005). What was found was novel and set the basis for many future hypotheses on metastasis. Those findings were that: 1) the cells persisted in the zebrafish embryos 2) the cells were motile and migrated to separate compartments that had developed due to tissue differentiation. This key precedent paved the way for future experiments to determine how human xenografted cancers will metastasize to different distinct environments albeit at a later stage of zebrafish development. What will remain to be answered, until now, is just how sparse these metastatic events are in addition to which regions or sub-compartments are systematically used in each biological or genetic test case.

In a pioneering paper, imaging analysis and statistics were used to determine the odds of finding a metastasis initiating cell in adult zebrafish. Injecting highly metastatic melanoma cells in optically transparent fish, it was determined that the odds of finding a metastasis initiating cell (MIC) was roughly 1 in every 120,000 cells or roughly 1/1000

xenotransplantations. These MICs were defined as a cell colonizing a site and creating a tumor distal from the primary injection site.

To register the fish to a similar coordinate system they used eye location as well as centroid measurements on N<20 fish for each cell implant cohort. From image registration, it was determined that there was a strong tropism for melanoma cells to congregate around similar cells in the zebrafish (Heilmann, Ratnakumar et al. 2015). However, there are key limitations on how this model would translate to an embryo xenografting model. These limitations are: 1) Adult fish (>90dpf) have an adaptive immune system, larvae do not, 2) larvae are injected with far fewer cells, but with less differentiated organs/site and 3) the statistical sample size is more than twice for larvae to allow for more statistical power. In comparison, zebrafish larvae offer a simpler biological 'filter' to show hotspots or areas of deposition of these rare metastatic lesions at higher statistical power. FishATLAS diffeomorphisms also have more defined local registration of the somites and other shape features. This diligence was required on the larvae level to achieve the aim of discovering metastatic hotspots. One of the first pieces of evidence for hotspots from a simple, yet pioneering composite adult fish registration algorithm is shown below in Figure 4.

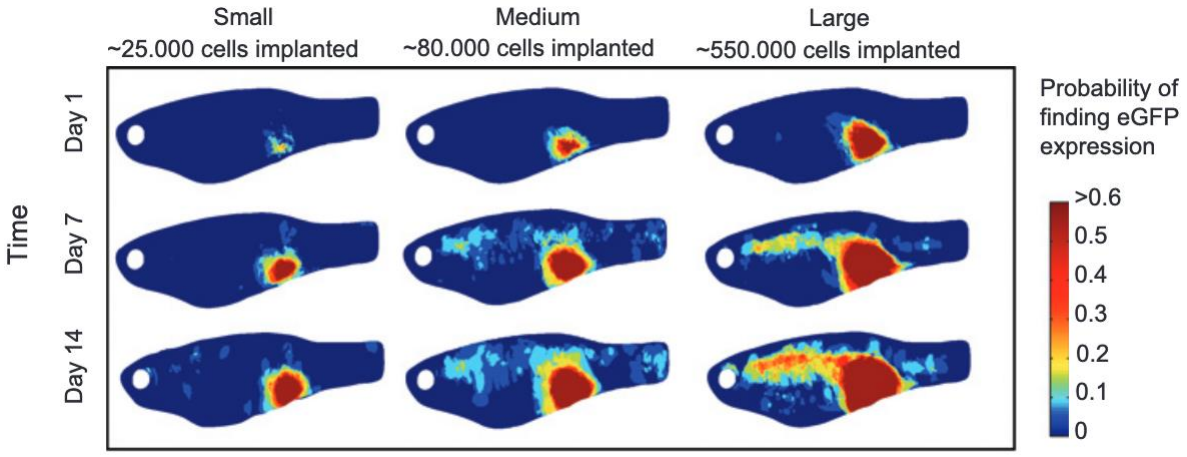



**Figure 3 Metastatic probability increases with cell implantation number in adult zebrafish. Pioneering zebrafish registration algorithm (N= 19, 20,14) for implantation of melanoma cells (Heilmann, Ratnakumar et al. 2015).**

This paper set foundational expectations for xenotransplantation studies that it is near-impossible to show hotspots without a rigid alignment of many fish or a very large number of cells. However, too many cells begin to raise questions of how biologically tractable it would be to inject a half million cells in an adult fish. To address this need in the field and make a pipeline for metastatic organotropism, FishATLAS was developed. The methodology of image registration, accumulation, image processing, and statistics are given in detail in the preceding chapter.

Xenografting experiments have progressed and exploded in popularity from a few papers in 2005 to having its all-time high in the past two years of nearly 148 papers on PubMed. What has changed during the progression was strongly reflective of the popular

technologies of the time. Genetic editing progressed from TALENs to CRISPR mediated activation (CRISPRa) and CRISPR knockout (KO), and from cancer cell lines to custom-made zebrafish cancer cell lines for melanoma (Heilmann, Ratnakumar et al. 2015). Each of these strides has helped to start establishing the zebrafish animal model as not only an alternative to murine studies but a rigorous standalone assay for metastasis.

FishATLAS develops on image registration algorithms to allow for the registration of hundreds of zebrafish xenografted larvae with high fidelity to visualize rare hotspot events by stacking each fish (called an accumulator). Each of these accumulations is orthogonalized to different conditional accumulators and their corresponding hotspots. Combining registration with conditional accumulators requires a statistical framework to determine:

1) Do hotspots exist in accumulation?
2) From two different conditional accumulators, are cell seeding distribution frequencies the same?
3) Are there shared sites of metastatic hotspots across many conditions?

These foundational hypothesis-guiding questions formed the aims for FishATLAS's contributions to the scientific community. From a literature review, it became clear that if progress was made on contributions between genetic and environmental pressure, this would offer a great tool for the scientific community.

## *1.5 Specific Aims of FishATLAS*

FishATLAS as a software package seeks to address the needs of the cancer research community by providing a robust framework of looking at both gene and microenvironmental pressure. The interplay of both environment and genes mirrors a central question of nature or nurture being a driving force in metastasis. To this end, FishATLAS addresses the following specific aims:

1) Determine the existence of metastatic hotspots for human zebrafish xenografts
    a) Functionally determine differences in hotspot deposition between EwS, melanoma, non-transformed cells, and EwS clonal expansions
2) Initiate investigating the role between intrinsic heterogeneity and microenvironmental pressure.
    a) Experiment on TC32 EwS monoclonal expansions to initiate investigations on clonal effects and environmental pressure
3) Determine conserved areas of metastasis between conditions to serve as targets of interest in cancer therapeutic studies
    a) Functionally map conserved areas onto biological landmarks

The structuring of these aims reflects research from a larger granularity of the existence of hotspots to smaller by looking directly at microenvironments of metastasis using transgenic markers for vasculature. Each of these aims has statistical testing regimens that are outlined in chapter 4 that help interpret these results from an imaging analysis or stochastic

method. An overview of each of these aims and a preview of their corresponding results are given below:

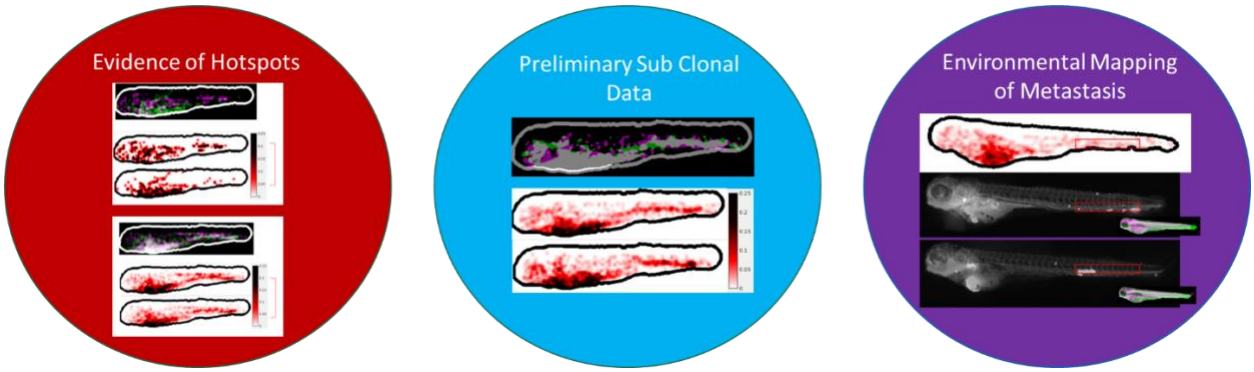



**Figure 4 FishATLAS determines the presence and environmental seeding of metastatic sites. Preliminary data with clones show statistical similarity between each clone.**

FishATLAS can prove the existence of hotspots leveraging its diffeomorphic registration at the heart of its methodology. All studies follow from using the high-fidelity registration to make inferences and statistical conclusions on metastatic tropism and distributions to allow for future guided goals in cancer drug discovery.

# CHAPTER TWO
# FishATLAS Methodology

## *2.1 Imaging Background: Xenotransplantation Imaging*

Zebrafish were chosen in large part for their optical imaging properties that are not found with any other vertebrate animal system. The larvae are optically transparent and lend themselves very well to putting transgenic cell labels or organism fluorescent labels via tissue-specific promoter sequences. This makes the regions that the cells inhabit and therefore, the microenvironment, defined in a much tighter method than can be resolved in a mouse model. The degree of uncertainly given by the lower resolution would be unable to meet the central goal of the project to determine specific organotropisms of cancer.

Stereoscope microscopy is commonly used as the primary imaging methodology for established xenografting laboratories. On occasion, for finer resolution studies of invasion or cell motility, confocal imaging is done in immobilized embryos, however, this has limitations for scan time (1-2 hours for confocal per whole fish) and phototoxicity. FishATLAS uses stereomicroscope data and Z-stacking for its diffeomorphic registration that is available for most laboratory uses. This offers great translatability for most labs to use this registration algorithm.

Typical analysis of zebrafish embryo xenotransplantation is done by taking a cohort of fish and selecting a few representative images from the dataset for visualization. One of the first papers on how metastasis is observed in fish was done by He et al in 2012 and showed different time points on the same fish and movement of individual cells down the caudal hematopoietic tissue (CHT). This data is shown below in Figure 5.

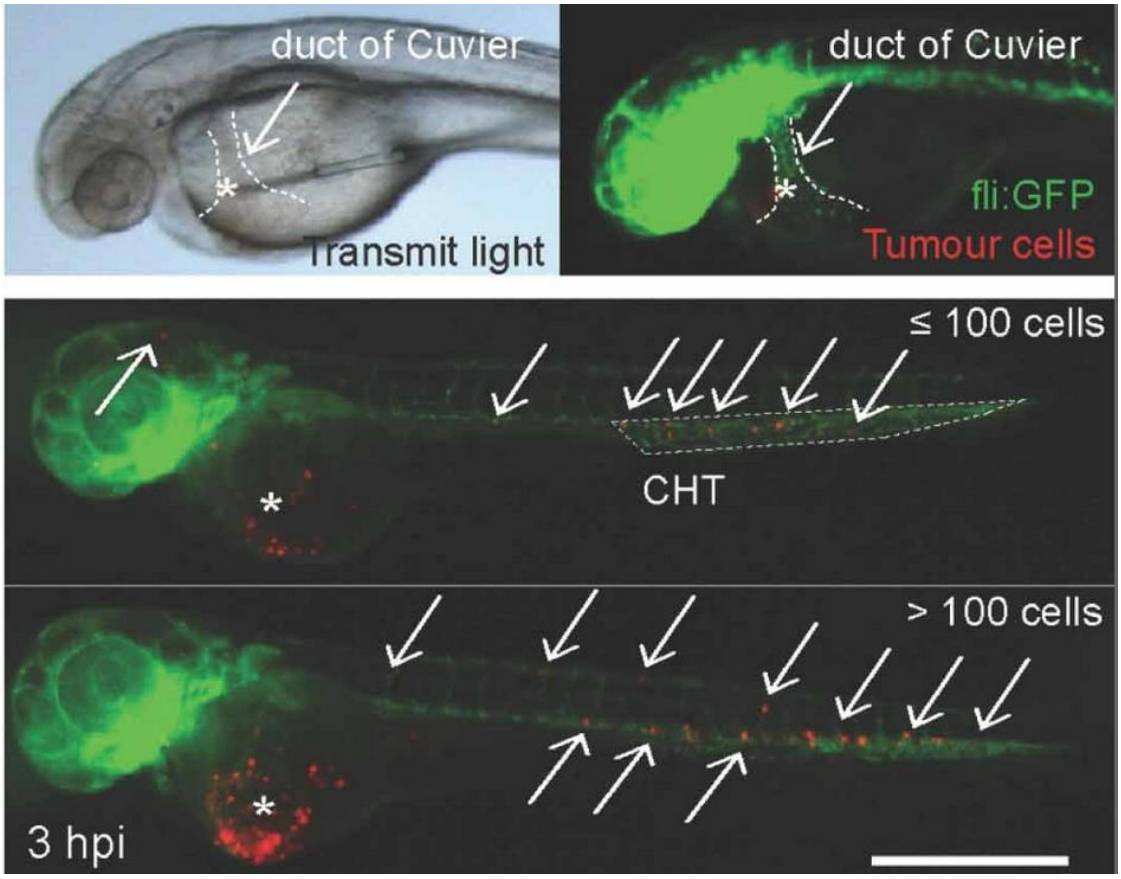



**Figure 5 Human cancer cell xenografts display metastatic behavior in zebrafish. Here injection site along the duct of Cuvier along with 0-hour post cell injection (hpi) and 3 hpi xenografting image (He, Lamers et al. 2012).**

This data was foundational in establishing how imaging could yield metastatic data and was later adapted to FishATLAS by taking numerous Z-stack images to allow for a full

focus image of the fish. Z-stacks are necessary during imaging to get each of the optical planes along the thickness of the zebrafish (roughly 3mm thick).

Data structuring for FishATLAS requires the imaging of brightfield (BF), green fluorescent protein band (GFP) with no cells, Red Fluorescent Protein band (RFP), and finally a depth map (DM) or binary mask showing the regions of the fish. To obtain these data for each fish, a Leica M107 Fluorescent stereoscope was used in addition to the LAS-X software for operations and Z-Stacks. To make computation more efficient, full-focus single images for each channel were generated for the final dataset in FishATLAS to serve as the final output.

### 2.1.1 Zebrafish Lifecycle

*Danio rerio* is truly an amazing species that offer a host of benefits for studying both diseases during development as well as adult diseases which commonly have been bourne with somatic mutations. At the initial single-cell phase, the zebrafish may be injected with transposon constructs to alter portions of its genome for scientific studies. One such transgenic fish is the Tg[fli:gfp] fish, where GFP is expressed under the same promotor region as the fli gene, commonly active in mesenchymal cells, cardiovascular system, and hematopoietic system (ZFIN 2021). This is just one illustration of injection of transposons at the single-cell phase and many other models of Cre-recombinase inducible systems as well as heat-shock protein systems are used. The zebrafish life cycle offers many different development steps over a short time and is given below in Figure 5.

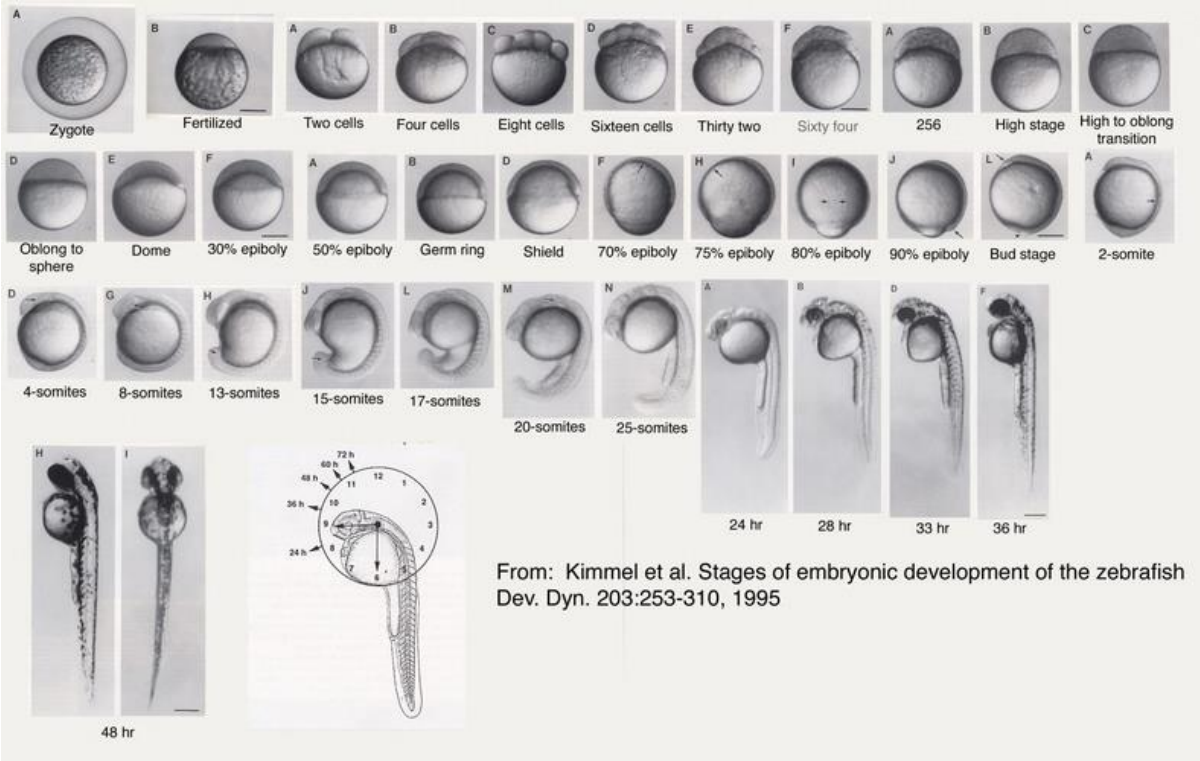



**Figure 5. Zebrafish's developmental life cycle allows for assays at many developmental time points. Xenografts are conducted at 48 hpf.**

Zebrafish are optically translucent for each of these stages and provide an attractive target for xenografting. Initially, xenografts were completed on blastula stage embryos (2.25 to 5.25 hpf) and used to observe cellular residence in different biological compartments across the zebrafish (Lee, Seftor et al. 2005). However, these early time stages lack a diversity of structures for microenvironment studies. In addition, at earlier time points before 48 hpf, the larvae are very sensitive to injection pressures due to the relatively soft outside tissue which results in a drop in xenograft survival. He et al. and other sources chose 48 hpf fish as an attractive, optimized timepoint for xenografts (Kohlberg 1978, He, Lamers et al. 2012, He, Lamers Gerda et al. 2012).

## *2.2 Zebrafish Xenografting*

Zebrafish xenografting has been a well-established protocol for metastasis assays(He, Lamers et al. 2012, Heilmann, Ratnakumar et al. 2015, Hill, Chen et al. 2018). However, some non-addressed nuance is cell-line dependent cell injection number, needle aperture shaping, the pressure of microinjector, and finally injection time. These are addressed below and will vary based on the equipment used, but the central principle of getting the right cell injection number remains the same. The following concise protocol is used for each genetic conditional cell line:

Zebrafish Mating Setup----------

1. Setup adult zebrafish 3 days before xenotransplantation
2. Collect embryos the next day in the midday or afternoon
3. Prepare cell culture plates for Ewing sarcoma cells 2 days pre-injection to ensure plate reaches 60-75% confluency on day of injection

Day of Injection----------

4. Prepare injection suspension by putting ~$2x10^6$ cells in a suspension of 1mL, spin down, aspirate supernatant and resuspend the $2x10^6$ cells in 20uL of 5% FBS in PBS solution. The goal is to get a working concentration of 50-60 cells per nL for Ewing sarcoma cells.
5. Using a petri dish with agarose, arrange 2 days post fertilization (dpf) zebrafish embryos on the plate and add 20-30 mL of E3 fish water + 2-3 mL of 0.2% Tricane for 7 minutes for anesthesia.

6. While waiting, prepare cell suspension by lightly flicking the tube and aspirate up capillary (Eppendorf Microloader) pipette tips and place suspension into a pre-cut capillary tube needle.
    a. On capillary needles, it is best to not use the typical pull shaping, but rather insert 1-2 manual delays to make a more ‘cone’ shape. This decreases the odds of cell clogging and lysis. Special care should be taken when opening the needle by using forceps lightly on the needle tip.
    b. The diameter of the needle opening is one of the biggest predictors for the number of cells injected (scales as $D^4$) so if a larger diameter is made, start with a very low PSI (0.8 psi) and iterate upwards from there.
7. On the lid of a petri dish, put a small drop of water on top and set the pressure of the microinjector (Warner Instruments PLI-100A) to a starting pressure of 1.8-2.5 psi and 0.10 seconds injection time. Depress microinjector pedal 2-3 times to prime cells in a capillary tube, and then prepare for count under transmitted light (TL) stereoscope.
8. Before injecting fish, optimize pressure setting to account for 80-120 cells per injection by counting the number of cells with each injection in a drop of water under TL stereoscope. Try to avoid storing cells in the needle upright, as this will result in more cells settling to the bottom and potentially clogging the needle, and getting an inaccurate cell injection number.
9. After 5-10 minutes, gently aspirate E3 +Tricane anesthetic solution off agarose petri dish.

10. Under the microscope, inject 100-120 cells into the perivitelline space of the zebrafish. Keep an eye for the cells displacing into the yolk through curdling-like motion around the yolk, this means that the pressure is too high, or the injection was done correctly. A correct injection will notice cells stay PVS and not disturb the yolk. Some cells might move down into the duct of Cuvier vasculature.
11. Repeat injection for subgroups of embryos and NOT the whole cohort as they will desiccate if left out too long on the dish. (When learning start with a group of 5-10, but more experienced injectors can do 20-40 as a subgroup).
12. Place all injected fish into labeled recovery dish with E3 and NO Tricane by rinsing them gently off the agarose injection plate.
13. Once all zebrafish are done, place in 34$^{o}$ C xenograft incubator (34C is the middle between 28-29$^{o}$ C for fish temperatures and 37$^{o}$ C for human cells)
14. 2 hours post-injection (hpi), screen fish for cells in vasculature and remove them from the primary incubator recovery dish.
15. 1 dpi, replace the fish water in the xenograft petri dish and remove all dead larvae.
16. At 2 dpi, euthanize fish in 0.1% Tricane/E3 solution and then place in 4% PFA in a cold room overnight
17. At 3 dpi, remove fish from the cold room and 4% PFA, rinse 2X with cold PBS, and suspend in PBS in a labeled glass vial until ready for imaging

    Imaging Day----------
18. Carefully remove and place individual fish on a coverslip with 4% Methylcellulose (spun down to remove air bubbles) for imaging.

19. Image fish by taking Z-stack images for each fish ensuring the entire plane of the fish is imaged.
20. Use Z-stack images to generate full focus images for zebrafish in the BF, GFP, RFP, and depth map (binary mask) space.
21. Save images and return fish to properly labeled glass vial with PBS.
22. Store fish in 4C in a closed dark box until needed again. Too much exposure to light will photobleach the remaining fluorophores in the cells. However, the intensity of cell fluorescent reporters will gradually decrease and be minimal or gone in roughly 4-7 months depending on fluorophore.

## *2.3 Diffeomorphic Image Registration*

Initial attempts in globalized intensity-based metrics using landmark centroids, rigid, and affine were met with difficulty in the region of the zebrafish tail. This presented a critical issue because the tail is one of the primary regions metastatic cells can go. Any registration error would discover hotspots impossible if there was no high-fidelity registration of fish somites and tail. An alternative was critical that focused on local registration and allowed for deformable shaping through an iterative method for the specific zebrafish morphology.

It is advantageous to computer science that the zebrafish has repeated shaping structures along the somites as well as a great amount of feature data from the eye. Indeed, in the critical tail region, even with biological variation from fish to fish, enough feature data was present to register local regions. The method used is the diffeomorphism, a local

morphing operator that relies on local gradients to smooth and iterate registration to a template to a high degree of accuracy (Chefd'Hotel, Hermosillo et al. 2002).

Diffeomorphic methods are used to estimate deformation maps to make a moving fish image fit a set template fish. There are many different ways to preprocess images, but what makes diffeomorphisms unique is their ability to handle image registration without the use of many images pre-processing (Chefd'Hotel, Hermosillo et al. 2002). In addition, it wouldn't require the use of the same imaging modality based on matching intensities. FishATLAS uses simple image processing to isolate the fish from background bubbles from the methylcellulose via a single pass entropy filter and binary mask to help with the fidelity of the registration via diffeomorphism.

Registration techniques require the calculation of a similarity metric such as the cross-correlation, correlation, or mutual information to set an endpoint for the algorithm. Each however can be computationally expensive and require a great deal of time especially in 3D images that FishATLAS might look to in the future. To address this issue, a method using optical flow was used to 'flow' diffeomorphic registration attempts that are driven by the smoothed mathematical gradient of the localized measure of similarity in a local region (ex. The tail, head, torso, etc.).

The next issue of registration is in biological reality: the variation of one fish's shape to another, which can be quite dramatic. It should be noted that all fish displaying non-

normal physiology post-injection are dismissed, but some fish persist with innate curvature along the dorsal fin requiring a larger deformation of registration to a template. These fish will be shown in a summary graphic at the end of this chapter in Appendix A, but constitute a realistic population of anywhere from 15-25% of fish imaging data.

### 2.3.1 Localized Similarity Calculations

Localized metrics of calculating similarity are varied. Consider two images: a moving image matrix ‘M’ and a template image matrix ‘T.’ The solution is to find a certain diffeomorphism D that convolved with M will match T within a certain tolerance. To begin we recognize an image is a collection of points, whose local point-wise joint density can be found locally over an image with dimensions IxJ by calculating the following:

**Equation 1**

$$p_{\phi}(i,j) = \frac{1}{\Sigma\, I(i,j)} \int G(T(i,j) - i, D(M) - j))\, dA$$

Where the calculation for the differences in diffeomorphism convolved image M, is the summation of all intensities, I, multiplied by the integration of the smooth Gaussian density kernel between the template and moving image undergoing diffeomorphic transformation.

Although a trivial definition, this approach can be adapted additionally by taking the above statistic calculation and normalizing it by the Gaussian signal G(T-I,D-j). For FishATLAS, this was determined to be unnecessary for the current implementation because there were no differences observed between normalized Gaussian point densities and non-normalized point densities.

Consider another metric for similarity, S(I, J) which looks at the pair-wise differences across a morphed moving image to a template using the joint density, $p_{\phi}$, the formula above. Using this, we derive a gradient of image similarity between moving and template fish, which is the optimization metric that is used for our stopping criteria.

**Equation 2**

$$\nabla S(\phi, i, j) = L(\big[(i, j), D(\phi(i, j), ij)\big)\nabla D\big(\phi(i, j)\big))$$

The L term in the above equation simplified is the local intensity difference between two different point densities creating a 2D gradient calculation. Despite localized intensity differences being used in the derivation of the diffeomorphic method, it has been proven in literature mathematically that these gradient evaluations converge to Gaussian smoothing of discrete histograms and are not dependent upon direct intensity measurements (Deriche 1990).

The gradient of similarity ($\nabla S$) is a key concept for this methodology. It is the hand that guides the localized morphing procedures to a high-fidelity registration between moving and template images through iterations to a deformation map D. This heuristic derivation is available as a reference, but the solution is shown below mathematically (Chefd'Hotel, Hermosillo et al. 2002):

**Equation 3. Set of diffeomorphism definitions and iteration parameters for deformation field.**

$$\phi_k = \phi_{k-1} + \epsilon R\big(\nabla S(\phi_{k-1})\big)$$

$$v_k \equiv L(T, D(M))\nabla(D(M) \equiv Displacement\ Field$$

$$\epsilon_k = R(v_k) \equiv Regularization\ of\ Displament\ Field$$

$$\phi_{k+1} \equiv \phi_k * (\phi_0 + e_k v_k) \equiv Diffeomorphism\ Iterative\ Variable$$

FishATLAS utilizes the diffeomorphism iterative variable ($\phi_k$) to continuously check and develop the displacement field to a certain similarity metric value $\nabla S$ with a given initial guess ($\phi_0$). Despite appearing computationally expensive, an individual fish can be registered in approximately 10 minutes per fish including all pre and post-processing steps. This might present challenges at a larger template and moving sets, but still stands as a great step forward for the community in registration for tumor microenvironmental studies.

Applying this method to three different fish in Figure 6 below, it becomes clear just how successful diffeomorphism is for the zebrafish xenografting community and cancer organotrophic studies as a whole.

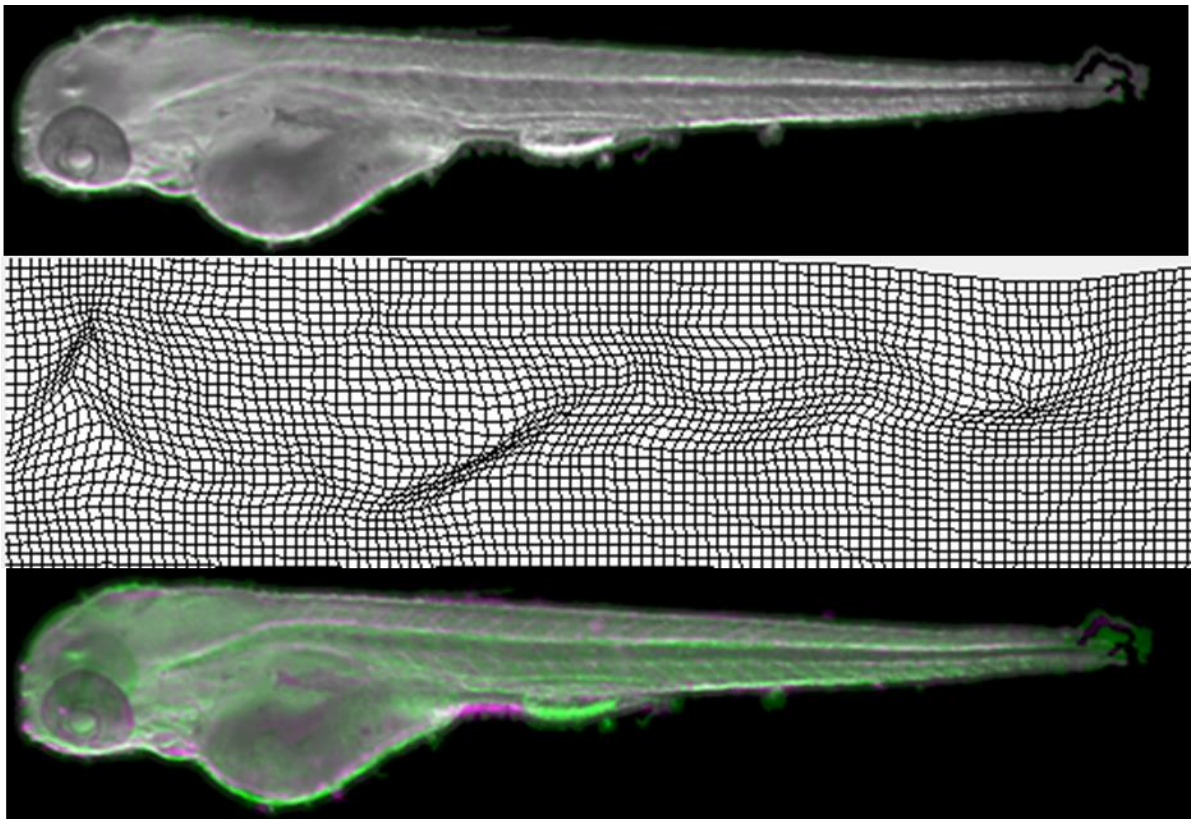


**Figure 6 Diffeomorphic registration of zebrafish registers regions with high fidelity making accumulation of cancer cells possible. (Template fish (top), diffeomorphic distortion map of moving fish image (bottom, magenta) to template (green).**

Each fish's registration is later evaluated as some registrations result in very extreme distortion fields. These are removed from the FishATLAS pipeline through histogram analysis of all mean square errors between morphed image and template, and removing the top 10% of outliers or fish that fall outside of 80% of the footprint of the template image. Over 8 different conditions, these post-processing controls have resulted in a well-registered cohort of fish accumulation data of over 1000 registered fish.

## *2.4 Cell Detection and Accumulation in Image Data Structures*

Cell data is obtained from fluorescent channel data (RFP), what remains post-registration is the detection, sorting, and accumulation. The current precedent in literature represents datasets that showcase a few representative images with a few going to whole dataset quantification of spread and imaging representation. Current standard methods are to do a rough image alignment and put all cell detection hits on a Cartesian grid. However, this neglects to address the difficulty of tail registration and biological variation of the zebrafish for the entire experimental cohort. Such an example is given below by Franzetti et al:

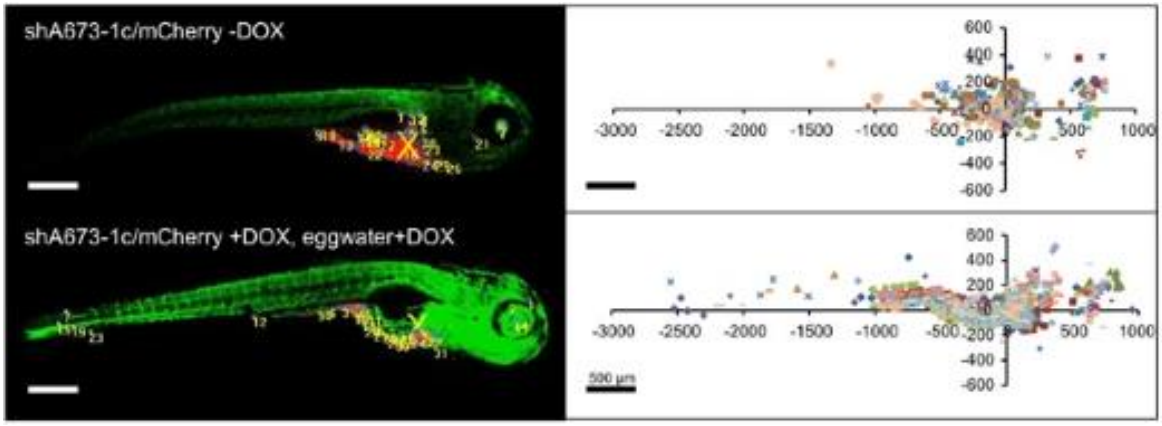



**Figure 7 Cell counting systems between fish have been attempted but lack rigid registration between each fish. Tail data and location can greatly change the locational mapping of cancer. (Franzetti, Laud-Duval et al. 2017).**

Cell detection in FishATLAS utilizes a Difference of Gaussian (DoG) filter to detect local maxima in the fluorescent channel data for each registered fish. DoG is a method that takes a fluorescent image and subtracts a larger and smaller gaussian signal to remove the larger, smeared regions of autofluorescence that may be present in the original fish image. A

representation of what DoG does and how it helps increase spherical signals of a certain size is given below in Figure 8.

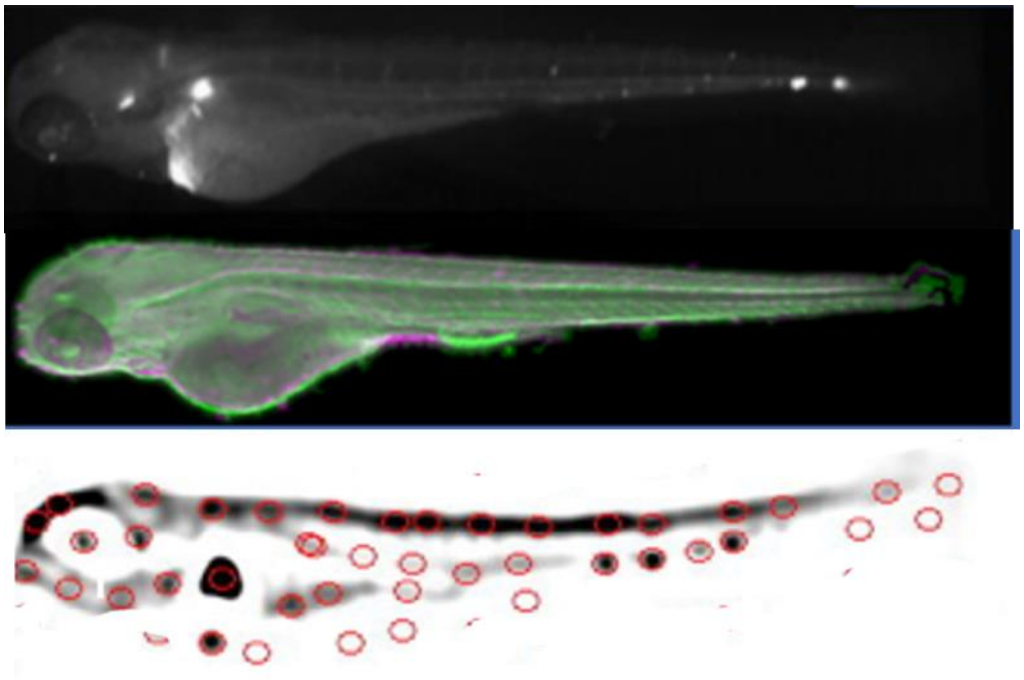


**Figure 8 Difference of Gaussian on diffeomorphism data makes detection of cell local maxima easier and allows for high-fidelity accumulation without background noise. (Top-Raw fluorescent image, Middle-Registered Fish(Magenta), Template Fish(Green), Bottom- Difference of Gaussian on Registered Fish Fluorescent Channel)**

DoG serves as a way to increase the contrast of spherical shapes within the band of the subtracted Gaussian signals, (sigma high = 10, sigma low = 2). These images are then readily detected using a local maxima search for pixels within a 5x5 neighborhood with the highest value (denoted above as red circles.) This helps remove the need for intensity-based thresholding in images and allows the same method to be used independently of the fluorescent protein. Intensity-based methods, when historically attempted, are very fickle and can have a different expression on a cell-to-cell basis. These DoG local maxima points are

then placed through one last filtration before being placed in the accumulation space. This methodology eliminates the need for contrast enhancement that could vary as a function of the fluorophore or experimental conditions.

Even with a large amount of signal-to-noise ratio (SNR) gain from DoG, there are still non-cell detection events. Not everything is removed by the DoG, so an additional step is to utilize a distance-based clustering algorithm of each of these points on a histogram. This method is illustrated below in Figure 9.

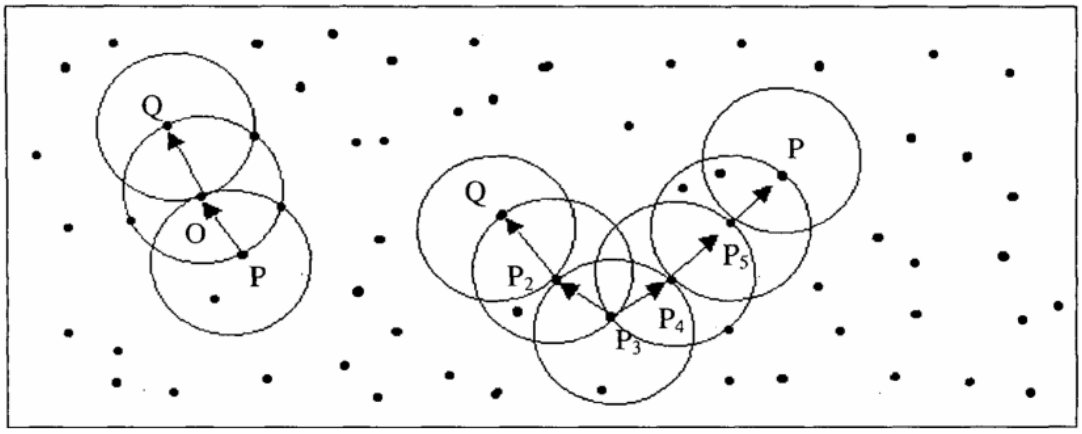



**Figure 9 Size-based plotting in DBSCAN shows areas of clustering that are based upon the sampled data population and independent of intensity.** Analysis of different points (P) and their corresponding density and reachability points (Q) (Borah and Bhattacharyya 2004).

Distance-based clustering statistics (DBSCAN) is a key implementation in the cell detection portion of this code to maintain some independence from intensity-based metrics. A key pitfall of some imaging analyses is the inability to account for changes in the distributional clustering of cells because of non-biologically mediated changed (ex. Different

scope, fluorophore, etc). This method relies on clustering together local maxima points that have many neighbors, which handily removes autofluorescence signals based on pure clustering patterns and demonstrates an ability to read the larger cell sites as well as the more faint ones seen in Figure 10 below. This again sidesteps any issue related to fluorescent channel or intensity-based metrics to allow usage for a wider range of applications.

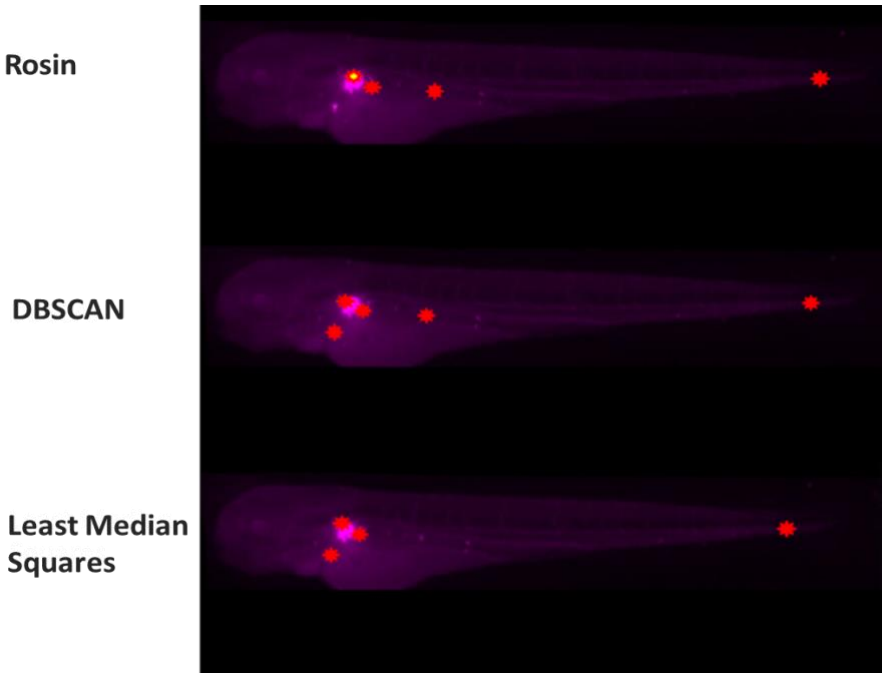



**Figure 10 DBSCAN shows more sensitivity post filtration of finding small and larger clustering sites in registered fish data.**

Once detection and registration have been completed, the accumulation spaces may be completed by creating a Gaussian signal over each of the detection points (shown above in red asterisks) for visualization of metastatic hotspots. These signals are normalized, and the units for the accumulation space, due to the particles being detected as local maxima, would be the percent of cells at a given location. An example of these accumulations between shEWSFLI1 and shCTRL is given below in Figure 11.

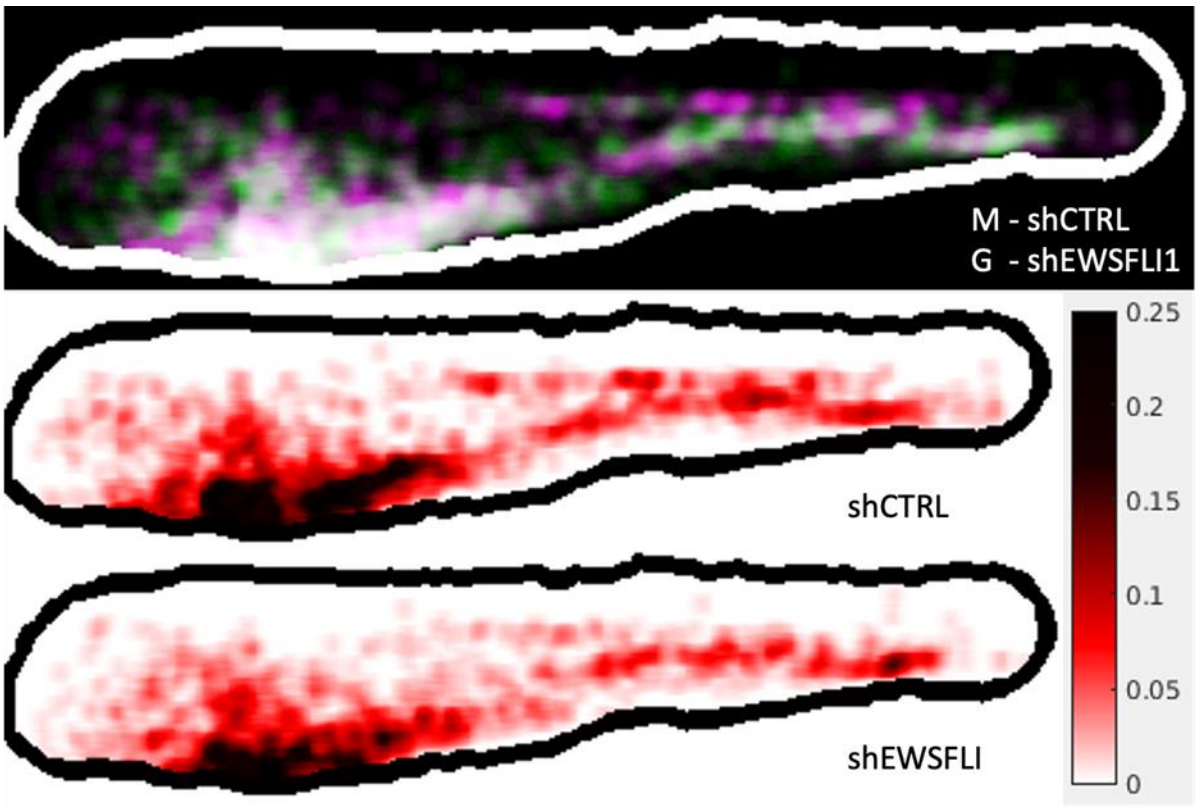



**Figure 11 Conditional accumulator of EWSFLI1 (shEWSFLI and shCTRL) illustrate the power of the registration and detection pipeline for determination of metastatic hotspots.** Composite Image (Top), conditional accumulators for shCTRL and shEWSFLI1 (Below), units are percent of cells at location.

understanding how difficult this process is and potentially why it has remained elusive to this degree in literature.

# CHAPTER THREE
# FishATLAS Accumulator Results

## *3.1 Bootstrapping Analysis: Determined Required Number of Fish for Statistical Power*

FishATLAS addresses the problem of statistical power in xenografting assays. In mice, obtaining a high sample number (N) can be costly and image modality is limited due to tissue opacity. In contrast, the zebrafish system facilitates experiments with higher N and thus is conducive to statistical rigor for xenografts. (Beyond statistics, there are also biological advantages as well discussed in the introductory material.) However, the sample size (number of fish needed to make a statistically significant conclusion from a given experiment) has not been established. The existing literature of zebrafish xenograft models describes sample sizes from n= 7 up to n=30 with the overall mean of these studies reaching roughly 20-25 fish. This poses a question as to how many fish reflect a full distribution of metastatic data?

The number of fish required for a well-sampled distribution was first taken by gathering a data cohort of N≥120 fish (EWSFLI1 test case) for both a scrambled shCTRL and shEWSFLI1 case. These data were put into a traditional bootstrapping method where a range of 1 to N random images have a test statistic calculated X times (to give an idea of variance in the data). This variance is expected to be quite large when low numbers of accumulators are included (low N) because of fish-to-fish variability. Statistical

bootstrapping is a very useful tool that allows for minimal assumptions of data distribution normality to find the required amount of samples to represent an unknown data distribution. A useful schematic of this is given below in Figures 12 and 13.

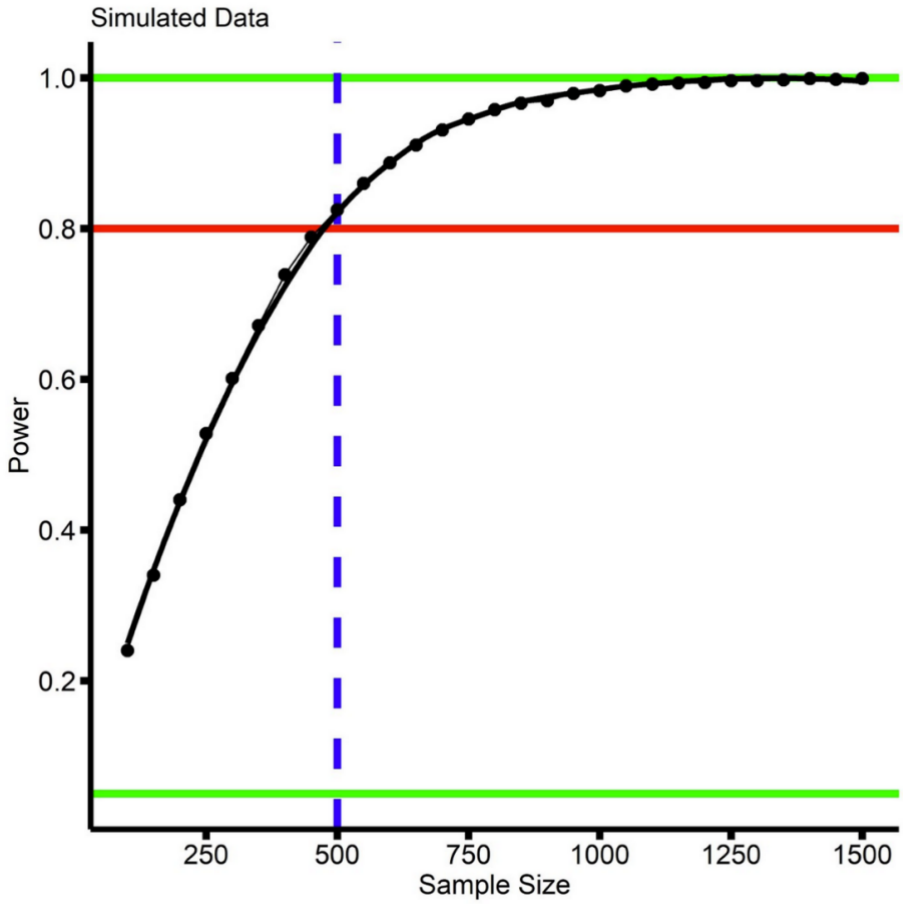



**Figure 12 Bootstrapping power analysis curve demonstrating how to determine statistical power as a function of sample size (mean values shown, statistical cutoffs shown at P=0.8).**

This process was repeated for each of the different test cases in FishATLAS to test whether specific experimental conditions or cell lineages influenced the required sample size.

Overall, these tests showed an ideal sample size ranging from 42 to 51 fish needed for ideal distribution.

At finer detail, to complete a bootstrapping analysis, there needs to be a chosen test statistic that encapsulates the hypothesis and question of interest. Metastatic microenvironmental location is the question of interest, so therefore, the Jaccard Coefficient was chosen, which looks at the area of the intersection divided by the union of two separate cell image masks (Equation 4).

**Equation 4**

$$J(A,B) = \frac{A \cup B}{A \cap B}$$

Bootstrapping was completed using the calculated Jaccard coefficients of two same-size, same-condition fish accumulators of size N, being resampled X=[1000,10000] times. These values generate a characteristic curve similar to the schematic above that begins to plateau at increasing N, suggesting a convergence to a standard sampling distribution that encapsulates the data. This bootstrapping analysis for shCTRL was done while thresholding the overlapping accumulators for 1, 2, and 3 cells at a given location is given below in Figure 14.

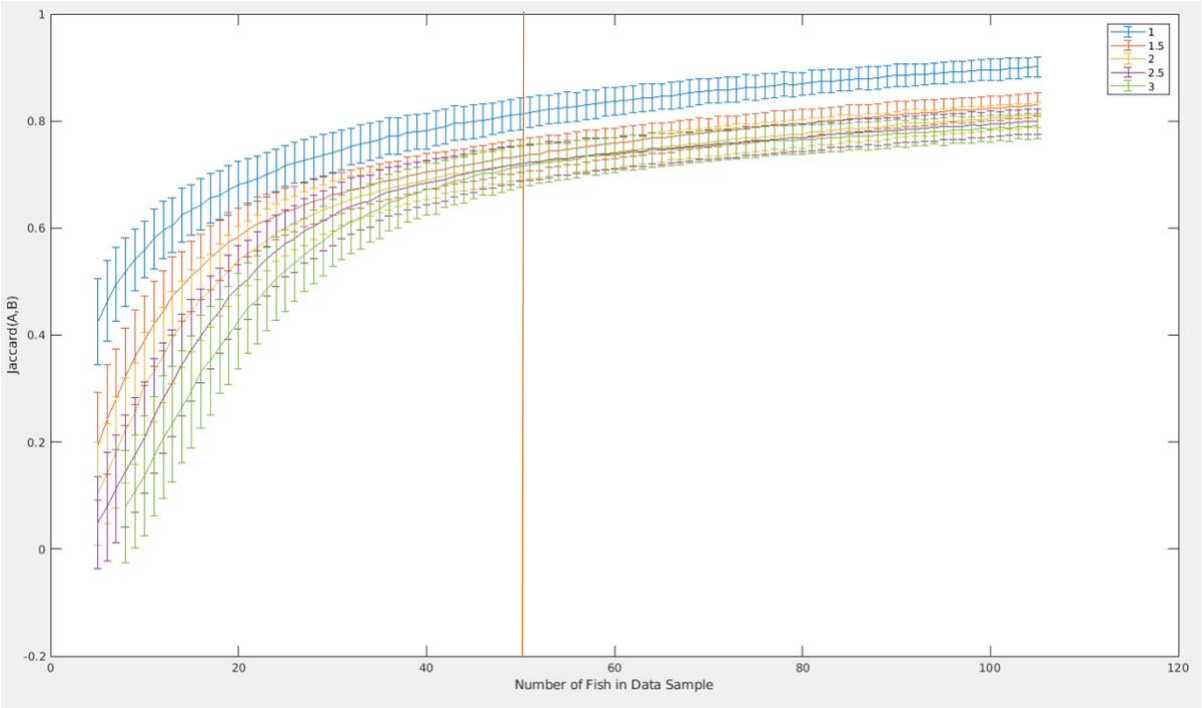



**Figure 13. Bootstrapping analysis of FishATLAS Accumulator Images for calculation cutoff thresholds of 1, 1.5, 2, 2.5, and 3 cells at a given residence site show an ideal selection of N=50 fish.** Different curves outline mask calculations at [1,1.5,2,2.5, 3] cell thresholds when calculating the Jaccard showing more and more specific site selection.

The above figure offers many crucial insights for sampling in the accumulator space. For one, it establishes the number of fish that are needed to obtain a good representative distribution to be roughly N=50 fish. The initial threshold for cell residence has a huge impact on sample size determination that greatly changes the shape of the curves themselves, which is to be expected as the assay increases stringency and selectivity of cell deposition sites.

For these studies, the sample size is typically chosen as the ‘elbow’ seen from the slope change at the outlined red line occurring at around N= 45-50. This data distribution can be further tested for optimal sample size by doing a local maxima test via the second derivative looking for f’’(x) = 0 (Equation 5) inflection points corresponding to an ‘elbow.’ This would indicate where the rate of the rate of change begins to show plateauing behavior.

**Equation 5**

$$\frac{d^2}{dx^2}(f(x_{inflection})) = 0$$

Applying this to Figure 14 below, the required number of fish necessary for an optimal test can be determined.

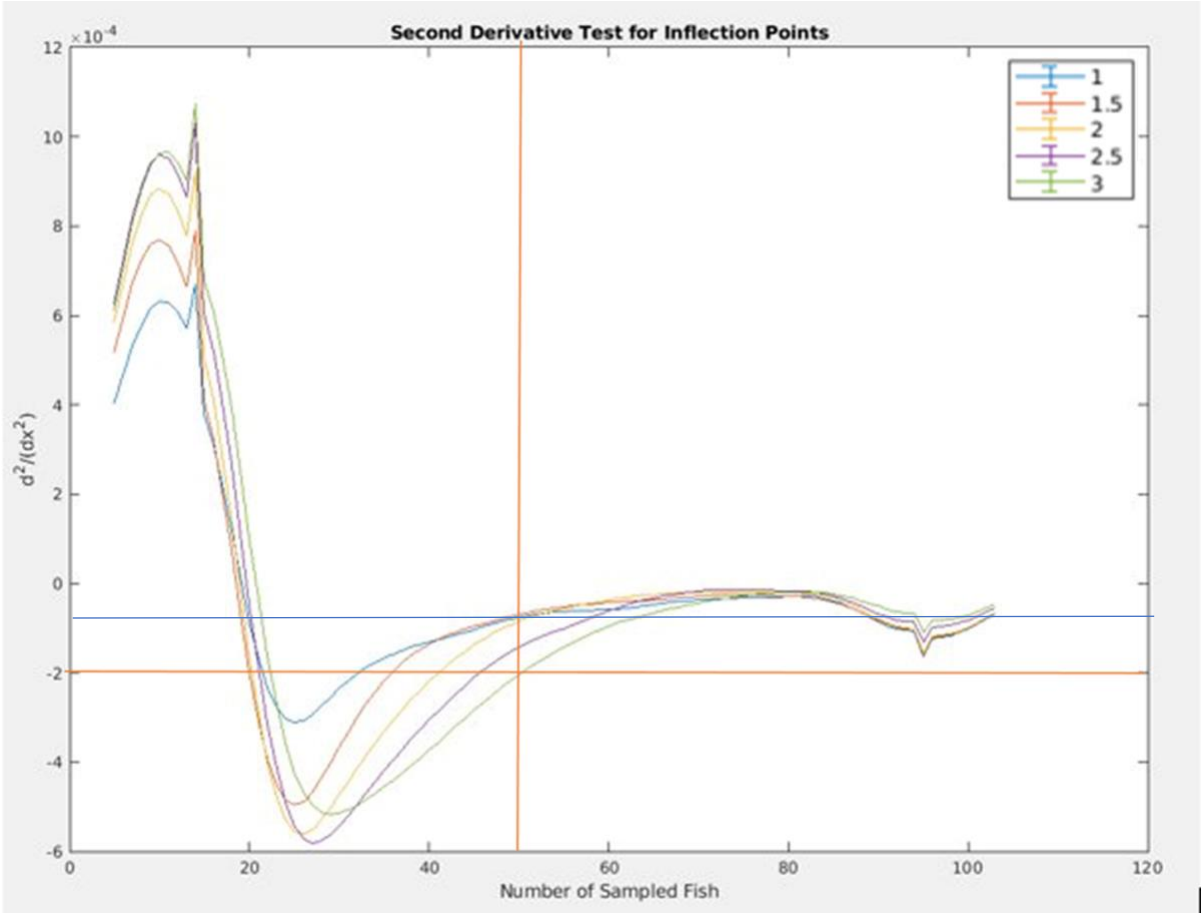



**Figure 14 Second derivative analysis of bootstrapping data showing the slope of zero (localized maxima) beginning at N=50 showing an optimal zone for sampling.** Different curves outline mask calculations at [1,1.5,2,2.5, 3] cell thresholds when calculating the Jaccard showing more and more specific site selection.

Based on the second derivative analysis, at most cell thresholds beginning at N=50, there is a clear flat leveling off as the rate of change remains stable, corresponding to the constant slope in the original bootstrapping plot. The artifacts seen at the end of the graph are artifacts due to sampling.

As a final step, this sampling selection may be shown by accumulating N zebrafish using FishATLAS. Each successive increase in sampling size demonstrates how each numbered step adds more and more detail to the total metastatic and non-metastatic cell distribution. To illustrate this, shCTRL fish with total sample size, N=117, were allowed to be sampled with replacement for N=[25,33,50,117] fish. Based on prior data, we should see that there would be minimal missing areas or hotspots for N>50 fish. These accumulators are shown below in Figure 15.

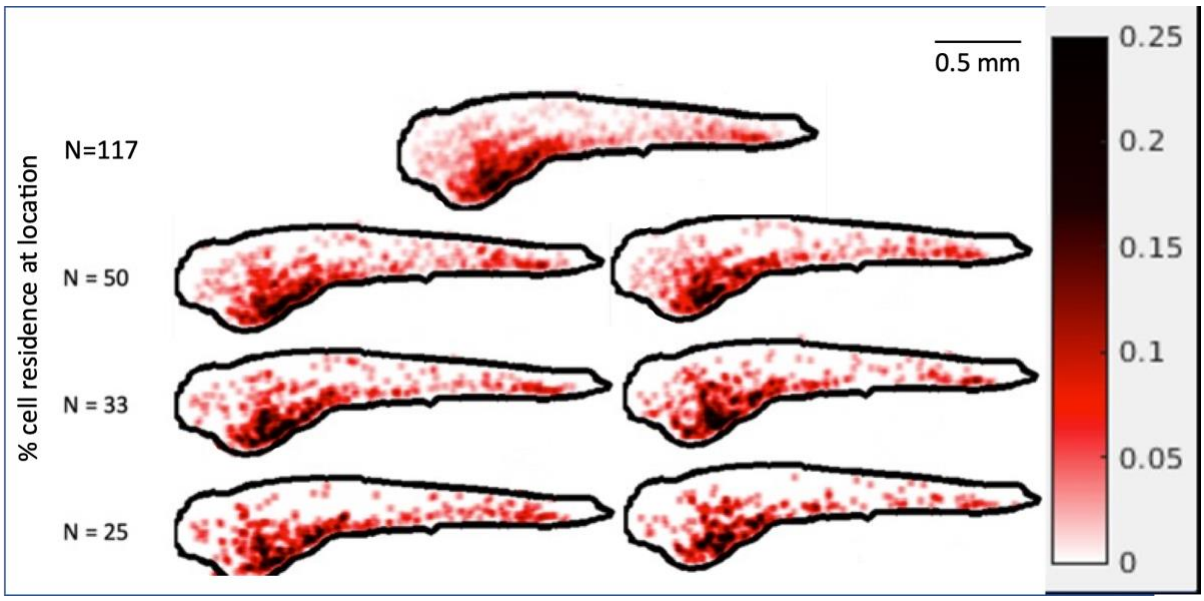



**Figure 15 Bootstrapped EWSFLI1 (shCTRL) accumulators for N= [25,33,50,117] fish showing the importance of N when using xenografting assays of sparse events like metastasis.**

It can be immediately appreciated that the calculated expectation of N= 50 performs very well visually in comparison to its lower N counterparts. Many of the hotspots consistently seen at levels of N>=50 are markedly absent at lower levels. Indeed, looking at the N=117 case, all sites are present and contain some degree of Gaussian noise around the finer puncta.

This result is to be expected as we are stacking nearly 2X the amount of Gaussian signals around these hotspots, so some addition of noise is normal.

In summary, this bootstrapping power analysis result is striking because many current publications **have less than 50%** of the required number to draw a statistical conclusion from zebrafish xenografts (Gaudenzi, Albertelli et al. 2017, Hill, Chen et al. 2018, Fazio, Ablain et al. 2020). Although many interesting findings have been gathered from xenograft literature, developments in sample size selection remain a crucial weakness to be addressed.

## *3.2 Non-Cancerous Cells: NIH 3T3 Fibroblast Accumulations*

As a negative control, non-neoplastic cells constitute a crucial FishATLAS accumulation condition. To this end, a well-established cell line for fibroblasts with a transgenic cell reporter in the RFP channel was necessary. These cells were injected at a standard concentration of $1x10^6$ cells/mL, the same as cancer cells, to allow for roughly 100-120 cells/xenograft.

Experimentally, something to note of fibroblasts is their completely different survivability compared to cancer. For one, cancer cells have shown robust presence at 2 days post-injection (dpi). Throughout this project, NIH3T3 cells show presence at 1dpi but seem to be completely cleared by 2dpi with very rare exceptions, making it nearly impossible to gather enough for a statistically viable population. This makes a comparison of the cells'

accumulators both difficult (cancer being evaluated at 2dpi) yet striking in the sense that only cancer cells have persisted longer *in vivo* from the cells that have been injected. A possible explanation is that there is higher fidelity immune clearance with the fibroblasts, however, no evidence was found to support this hypothesis other than complete clearance of fibroblasts at 24 hpi. Although lacking an adaptive immune system at the time of engraftment, zebrafish contain a fairly active innate immune system that has been shown in previous data to interact with cancer metastases or other foreign tissues (Tulotta, He et al. 2016).

This non-cancer accumulator was completed on a dataset of N=53 fish, corresponding to a total accepted count of xenografts, not the total amount of image data. This accumulator is given below for fish seen at 1 dpi.


**Figure 16 Fibroblast (NIH3T3) accumulator of N= 53 fibroblast cells showing very limited specificity with regards to cell deposition or location.** Units indicate the percentage of total fish population with a cell in a given region or hotspot.

Fibroblasts show little movement from the injection site. This was an interesting result as even with the physical environment of venous pressure and blood circulation most

NIH3T3 cells were unmoved from primary injection space. To be complete, there is still what appears to be a singular hotspot at a similar region of the tail that is shared with cancer. To look at this further, we look at a side-by-side comparison of accumulators in regions of greater than 1, 2 and 3 cell occupancy. This is shown below.

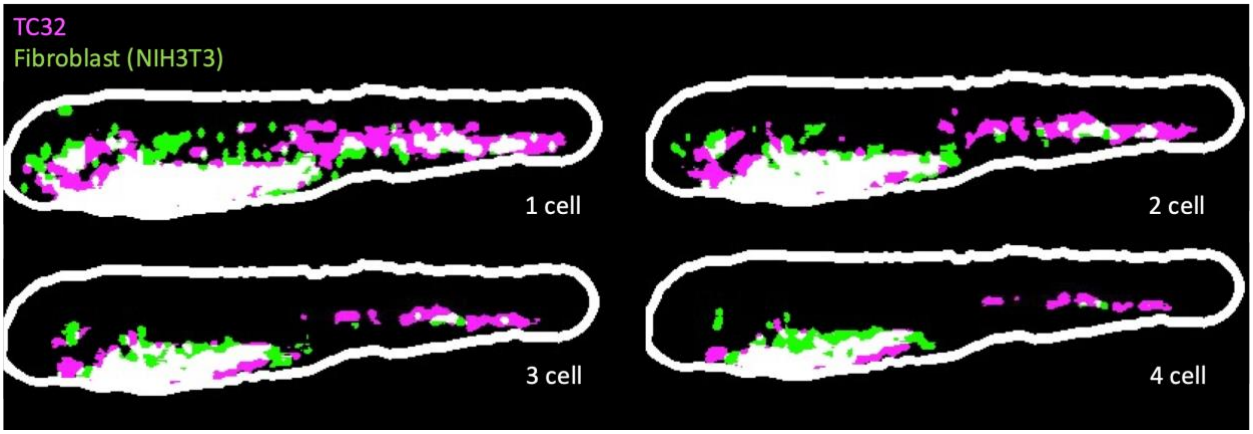



**Figure 17. TC32 Ewing sarcoma cells persist in their hotspots at more stringent cell thresholds than NIH3T3 fibroblasts.**

At higher levels of cell occupation, hotspots in fibroblasts differ from what is observed in cancer cell lines viz. disappearing at a 3 cell residency threshold. The presence of puncta or specific regions is very clear upon accumulation, but this stresses the need look at this complex phenomenon in a cell residency manner. In the proceeding chapter, differences in cancer versus non-cancer is presented in the statistical results chapter in the form of a p-value, with additional functionality to show which region/s are shared between fibroblasts and all other cancer with or without genetic modulations.

To summarize, NIH 3T3 fibroblast cells show a pattern of cell deposition that shows weaker puncta or specificity in comparison with what is seen with cancer cells. The

deposition from the injection site down to the CHT. Each of these findings is in stark contrast to what is seen with both melanoma cancer cells, TC32 shRNA constructs, and TC32 clonal expansions showing the inherent accumulator differences between cancer and non-cancerous cells.

## *3.1 Melanoma Cell Lines: Metastatic MV3 and A375 Accumulations*

To determine how cancer cell accumulators differ on the level of cell type, melanoma, a very prevalent and well-researched cancer, was chosen as a comparison to future experiments looking at EwS. For melanoma experimental accumulators, MV3 and A375 cells were injected at a concentration of $1x10^6$ cells/mL in 2dpf zebrafish embryos. Imaging was completed at 2 dpi (4dpf), with images processed, registered, and accumulated in FishATLAS. These data are given below in Figure 18.

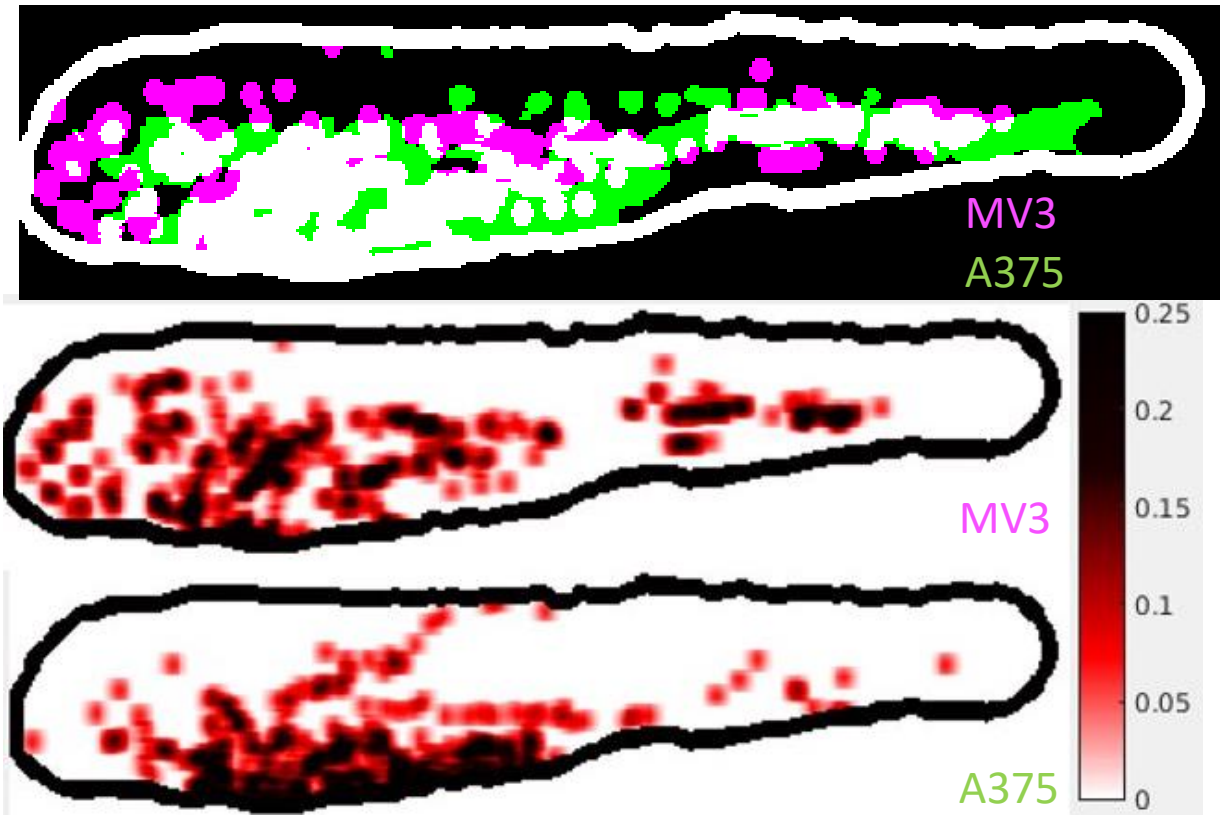



**Figure 18 Melanoma cell lines display different drastically different spreading patterns between A375 and MV3. (MV3, N=42; A375, N=63)**

From the melanoma accumulators, it would appear that the MV3 cells have high specificity in regions in the inter-somitic vasculature, while the A375 cells remain mostly localized around the site of injection and the yolk. Although these results demonstrate a stark difference, this is just one piece of the overall puzzle, and realistically to determine more about metastatic propensity would require more forward genetic modulation studies in FishATLAS as well as a screening of variants such as the Rac P29S mutation in A375 cells shown to increase metastatic ability greatly (Mohan, Dean et al. 2019). Friedl et al. hypothesize that MV3 gains metastatic propensity due to loss of ECM adhesion, based on in vitro studies in collagen. In contrast, work by Mohan et al. offers support that P29S mutations in A375 cells mediate a much higher contribution to invasion and metastasis. The data from this assay shows in this experimental setup that A375 and MV3 are significantly

different globally, and MV3 colonizes more hotspots from the injection site. Specific reasoning as to why this is the case is a future question.

An unplanned, yet appealing result in choosing A375 and MV3 is in the difference in mutation in a very established cancer pathway: b-Raf/Ras, (with A375 having a b-Raf mutation and MV3 neither). How this mutation affects metastatic conditioning/accumulation remains unclear. Indeed, this very result of the accumulation showing MV3 yielding increased metastatic hotspots remains controversial, as other work indicates through *in vivo* studies of PDX trained classifiers that A375 would be more metastatic over MV3 (Zaritsky, Jamieson et al. 2021). Although potentially a result of either side not encapsulating the full complexity of this issue, it gives to support the necessity of both *in silico* and *in vivo* being done in tandem to obtain a clearer picture of metastasis not being a one-axis, one-cause event.

It is well established in the literature that there is a great degree of heterogeneity from one cancer cell line to another, and this is seen in the melanoma FishATLAS accumulator data. These accumulators are the first of their kind showing hotspot coordinate differences on a rigid coordinate system. This finding can be used for drug discovery for any melanoma cancer to see changes in drug response for each microenvironment. As an example, a review of the literature shows many current papers are using B-RAF/RAS inhibitors. Although promising, historically this has failed in translation to the clinic as b-Raf inhibitors are plagued with drug resistance issues requiring other combination therapies that could cause

the potential for side-effects (Kim and Cohen 2016). FishATLAS assists the bench-to-bed-side process by allowing wide-spread assays of genes and combinations of drugs to be scored on not just tumor volume, but rather on the persistence of cells in metastatic-allied microenvironments.

## *3.3 SOX6: EWSFLI1 shRNA and shCTRL Accumulations*

*SOX6*, a gene of interest, is a direct target of the EWS-FLI1 transcription factor. Previously, Marchetto et al. that patients with elevated *SOX6* were linked with poorer outcomes with many having metastatic disease (Criscuolo, Marchetto et al. 2020). Precursor experiments to FishATLAS consisted of a simple cell spreading assay for a targeted shRNA for SOX6 and scrambled shCTRL. A schematic of this assay is outlined below in Figure 19 and the results of this assay for SOX6 on two different EwS cell lines: TC32 (SOX6 higher) and RDES (SOX6 lower) in Figure 20.

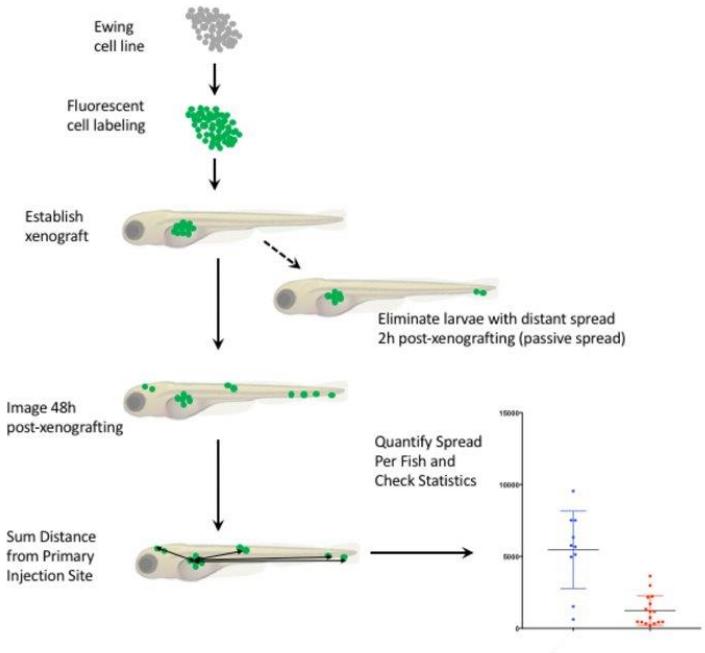



**Figure 19 Initial *SOX6* spreading assay used to determine the differences in cell spread from the injection site, where a single data point is the summed distance from all cells in a single fish.**

Each data point in the resulting graph at the end of schematic represents the integrated distances of all cells from the original injection site above the yolk sac in the PVS. Approximately 20-30 fish were obtained, and data points integrated for TC32 and RDES shCTRL and shSOX6 conditions. These results were then tested for statistical significance from scrambled shCTRL by use of a Kolmogorov-Smirnoff (KS) test comparing two cumulative distributions. This test was chosen for its insensitivity to non-normalized data (this data was determined to be log-normal). This data is presented in Figure 20 below.

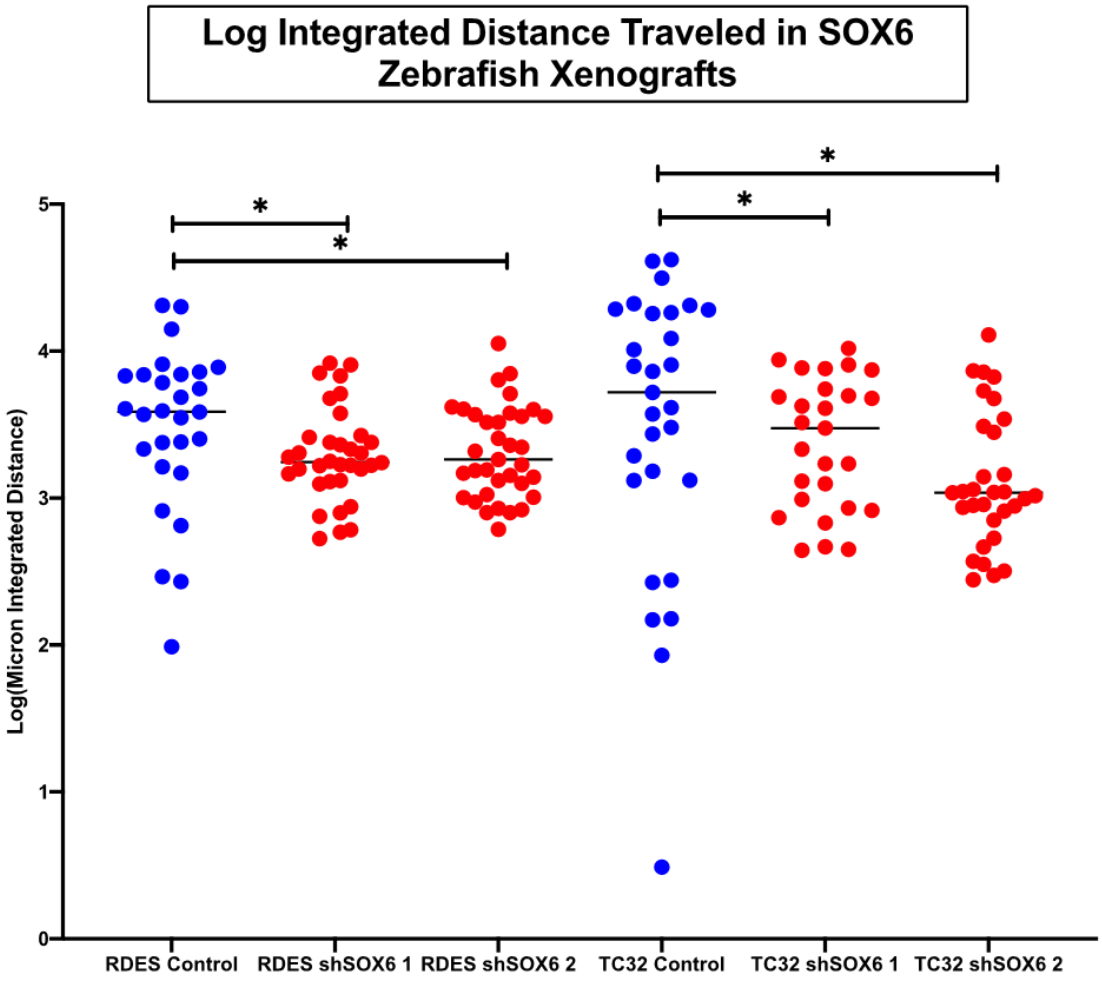



**Figure 20 Log integrated distance of TC32 and RDES EwS cells with 2 shRNAs and 1 scrambled shControl. (Significance determined from KS test, * = p<0.05)**

Cell spreading integration per fish shows a very striking phenotype of diminished cell spread upon *SOX6* shKD for both TC32 and RDES. The KS test between both shRNAs and scrambled controls (shCTRL) showed statistical significance in both cell lines suggesting that SOX6 has a substantial role in increasing the amount of cell spread from an injection site.

However, these data beget a key weakness: location. Without location, we are bereft of any microenvironmental information or potential for organotropism assays. This was the inspiration to use the diffeomorphic image registration outlined in the prior methodology. Utilizing this registration, the details of SOX6 become clearer on the regions of interest, culminating in the completion of one of the primary aims: to determine if and where regions of high visitation exist. Such regions are referred to hereafter as hotspots (HS). Figure 20 below is the accumulation of N=80 and N=113 fish for shSOX6 and shCTRL respectively.

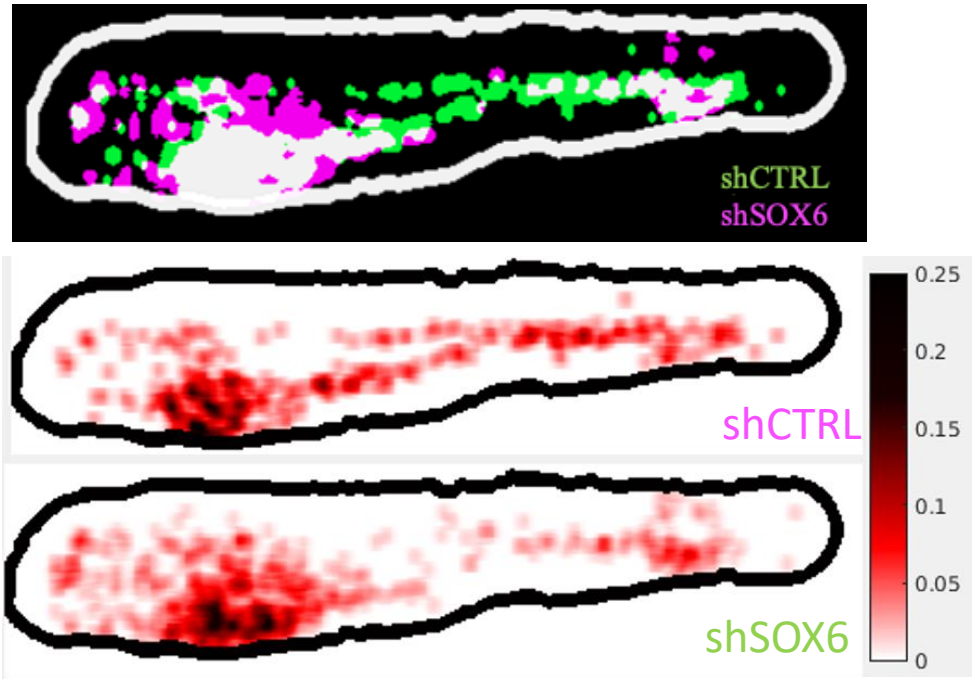



**Figure 21. SOX6 knockdown displays marked depletion in the central somitic region.**

These data show the presence of not one, but rather multiple hotspots. These results indicate for the first time, on a registered template with high accuracy, where these cells over many sparse metastatic events are located. Of particular interest are the regions of the head, caudal hematopoietic tissue (CHT) and dorsal vasculature/somite region. One particular item to call attention to is the is the almost complete absence of shSOX6 cells in the central somite region of the fish, as outlined in Figure 21.

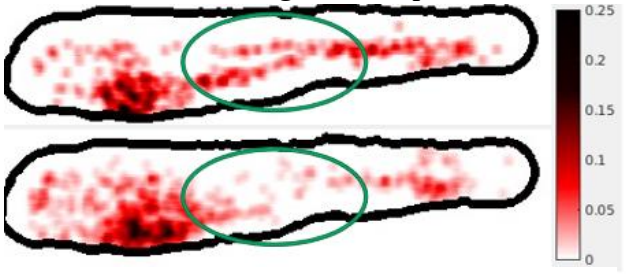



**Figure 22. Hotspots disappear in central somitic region upon shSOX6 KD.** Top- shCTRL, Bottom- shSOX6

Review of the literature and a previous collaboration have unveiled some potential mechanisms of interest on the SOX6 axis. Collaborators in our previous publication outlined the role of oxidative stress (or reactive oxygen species (ROS)) levels with SOX6. Their work stated that higher levels of ROS associated with SOX6 activity presents an exploitable drug target by increasing the already high ROS to a lethal level for the cancer. Elesclomol is a proposed drug therapy that works to exploit the elevated ROS and provide a more targeted chemotherapeutic (Marchetto, Ohmura et al. 2020). Merging these conclusions with FishATLAS, the data hints that SOX6 utilizes a specific mechanism that increases the internal ROS of the TC32 EwS cell itself, and these cells persist in more diverse environments than when SOX6 is under KD. Ultimately, what remains unknown as to how this would confer specific survival to cells in this central somite boundary. However,

knowing that SOX6 directly has an established role in modulating ROS is a good place to start a future analysis.

Preliminary experiments in determining the landscape of oxidative stress have shown potential evidence that the central somite space is one of differential oxidative stress in comparison to other tissues (Mourabit, Fitzgerald et al. 2019). Transgenic fish lines have been developed that could offer insight into microenvironmental cues guiding organotropism. An example of these reporter lines along with the representative accumulators for shSOX6 and shCTRL are given in Figure 22. Repeating this assay in WT AB/TL zebrafish with both hydrogen peroxide treated (positive control) and uninjected fish with MitoSOX Red, it was not determinable if the central somatic region presents itself as a potentially higher region of oxidative stress (Figure 23).

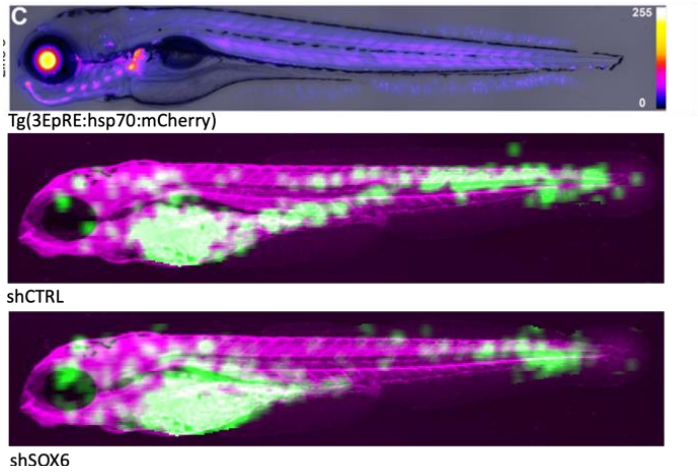



**Figure 23. Oxidative stress reporter suggest higher levels of oxidative stress in regions of SOX6 WT expression.**

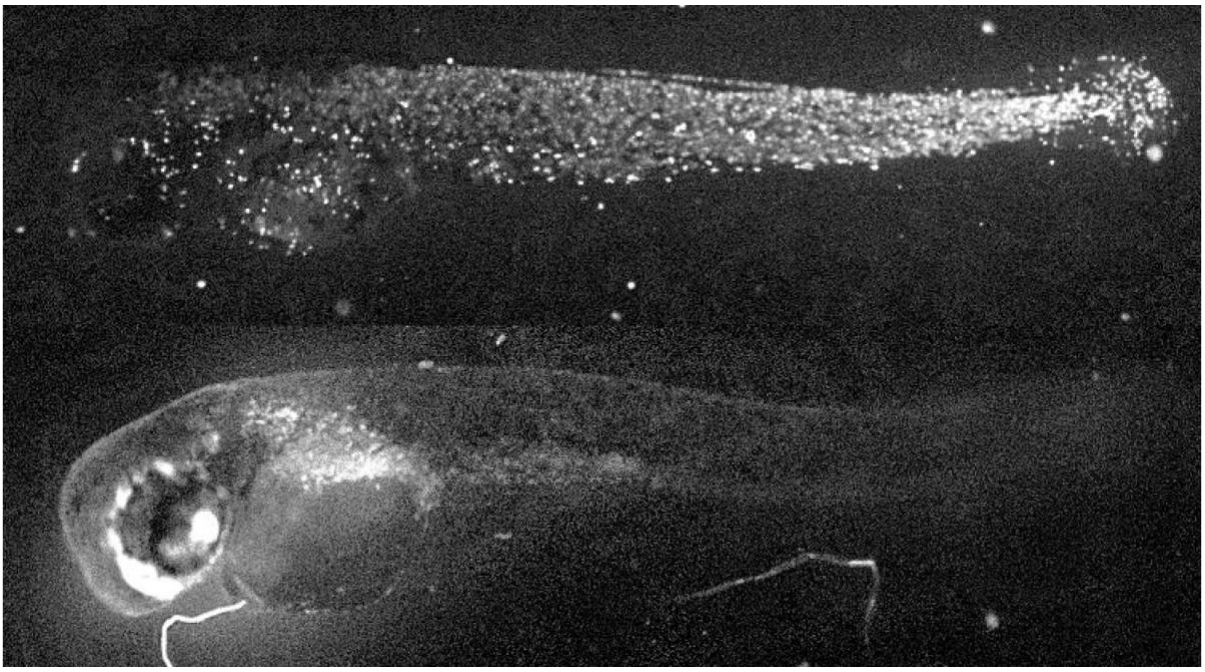


**Figure 24. MitoSOXRed oxidative stress staining for positive peroxide control (top) and uninjected fish (bottom) at 2 dpf.**

Attempting this using another fluorophore, CellROX green with an established protocol for embryos, unfortunately, was not even able to give a conclusive positive control response. This suggests future experiments requiring a more direct tissue and cell reporter to determine the interplay between both as oxidative dyes were unsuccessful.

The results of the MitoSOX red have some promise, but this finding needs to be further substantiated by looking at orthogonal methods such as an oxygen-sensitive biosensor in the Ewing sarcoma cells. A potential complication also arises at the protocol injection timepoint at 2dpf in the fish larvae cardiac space where primitive embryonic erythrocytes experience apoptosis around 2-3 dpf. This appears to present itself in Figure 23 in the upper

yolk spaces. To determine if this issue was going to eliminate this method, at the time of FishATLAS imaging (4dpf), this protocol was repeated, but no presence of oxidative stress was detected on the WT fish or positive control, suggesting some type of permeabilization issues with older embryos. This substantiated the clear need for alternative assays of oxidative stress and other environmental factors.

### 3.3.1 Sub-Accumulation of SOX6 Data

In many cases, cancer metastases are seeded in primary, secondary, and potentially more regions. This important spreading phenotype is something that can be determined using sub-accumulation. Sub-accumulation looks at accumulator cell distribution sites that share a hotspot of interest with a given primary region of interest (ROI), hereafter referred to as secondary sites. FishATLAS supplies the necessary workflow to determine not only where these locations are, but also the individual data (single fish images) that contributed to these secondary spaces. This is completed by taking, in this case, the SOX6 accumulator and allowing the user to draw an ROI from the accumulator to indicate the primary hotspot. FishATLAS will then accumulate all fish that had a cell in this ROI, leading to a secondary accumulator to determine if there are any secondary sites present. This step is incredibly important as it adds a level of hotspot contrast that might be invisible at the whole N accumulation. An example of this ROI sub accumulation is given below in Figure 24.

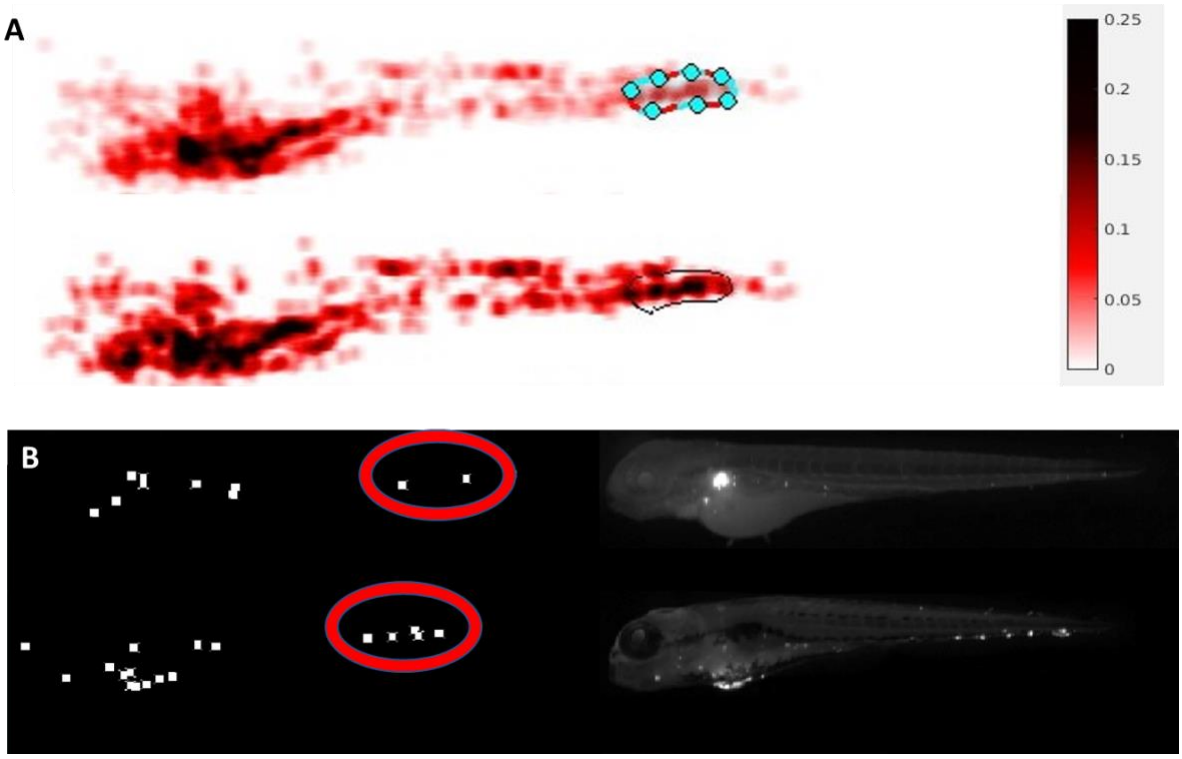



**Figure 25. Accumulator CHT ROI subaccumulation displays potential secondary sites of metastasis. A) Whole conditional accumulator (top, N=117 fish), subaccumulator of all fish with contributions to ROI (bottom) B) Representative single fish in ROI. Units: Frequency of cell deposition at location**

Upon sub-accumulation, secondary sites become much more pronounced. In the above, the dorsal venous regions, as well as the hindbrain, yield themselves as secondary hot spots of interest that are shared with the caudal hematopoietic tissue (CHT) ROI. These key points posture the accumulation system well for future drug studies that not only look at

globalized changes but how coupled primary and secondary sites respond to cancer drug therapies.

### 3.3.2 Mapping of Subaccumulators of Single Fish to Fish Morphology

Single fish data of interest from the accumulator space may be mapped to any existing template with morphological markers of interest for delineating TME. For the images above, it is hard to determine exactly where each cell falls in the singular fish data. To address this, FishATLAS registration works from both Fish Data to Accumulator and then from Accumulator Single Fish to Single Fish Template. Using the ROI in figure 24, single fish were registered back to a vasculature labeled fish (fli:GFP) and displayed below:

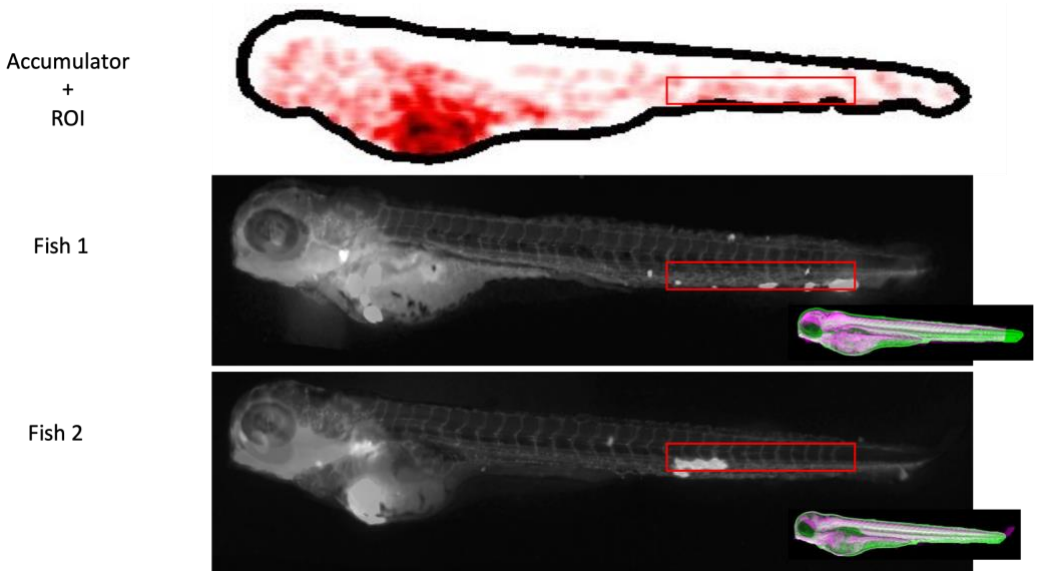



**Figure 24-B Environmental mapping of single fish from accumulator ROI shows high fidelity landmarking of metastases in vasculature and CHT.**

This is an incredibly powerful tool for determination at a higher resolution where single fish data lie. As the accumulation space primarily functions to show metastasis, this

functionality shows specifically where with fine resolution in a very large dataset of N>100 fish.

## *3.4 EWSFLI1 shCTRL and shRNA Accumulations*

The salient cause of Ewing sarcoma poses a key experimental outcome for FishATLAS. Accumulation on its transcription factor target, SOX6 yielded a strong depletion of cells in the central somite space. It was hypothesized that EWSFLI1 shRNA KD would demonstrate a similar phenotype.

In preparation for this experiment, cells stably expressing a doxycycline-inducible shRNA targeting FLI1 in the TC32 background were used. A parallel control shRNA targets a region of FLI1 that is part of the fusion oncogene and is not expressed in Ewing sarcoma cells and makes use of a DOX inducible mCherry reporter. RFP was chosen for its ability to not be strongly affected by autofluorescence caused by fixation in 4% PFA. This allows all fish to be preserved and reimaged as needed for future workflows.

The accumulation was completed for N>100 fish for both shCTRL and shEWSFLI1 and is displayed below in Figure 25. These fish constituted the primary data set for the bootstrapping power analysis that determined that a minimum of 45-50 fish are needed to give a representative distribution.

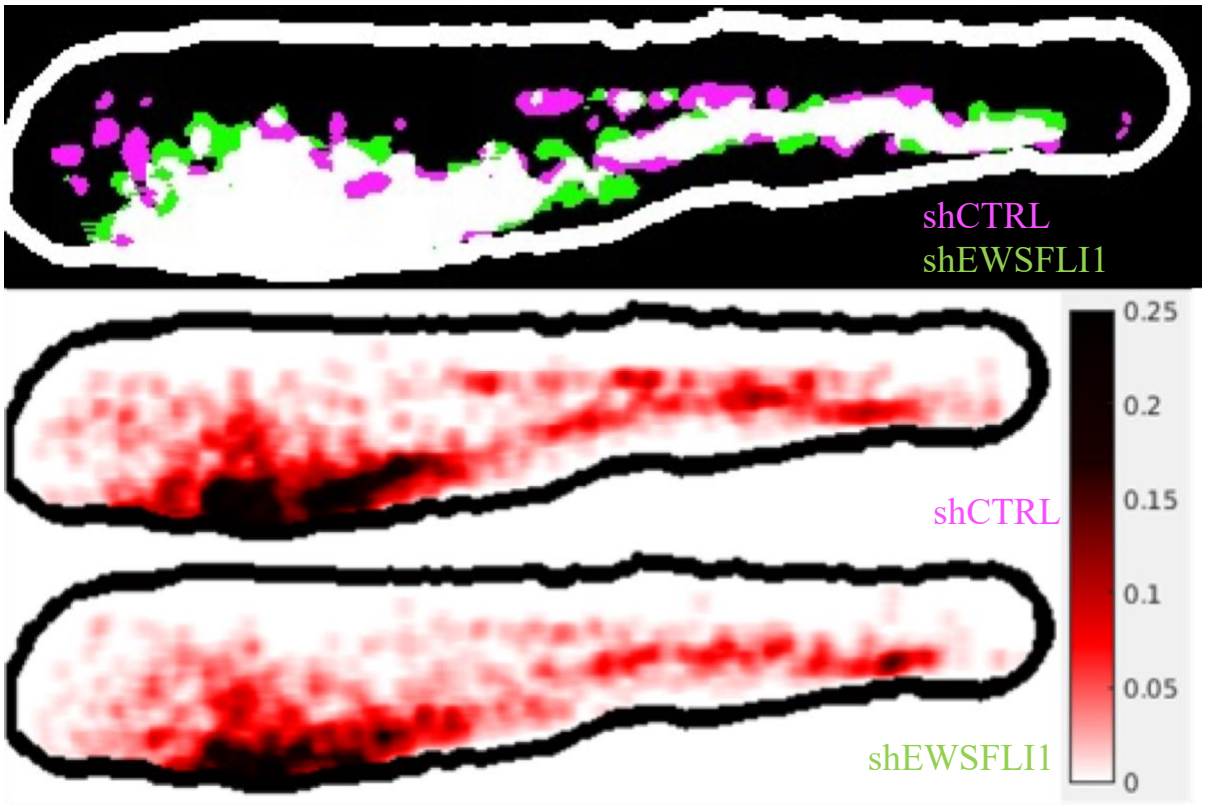



**Figure 26. EWSFLI1 shRNA-mediated knockdown displays subtle differences between deposition in the upper dorsal veinous tissue.**

Interestingly, a direct target of EWSFLI1, SOX6, presented an admittedly much weaker but similar pattern of cell deposition removal in the central somite/dorsal venous tissue space. However, the effect appears to be less dominant in removing cells from accumulating in this space, potentially demonstrating a more penetrant effect on the level of SOX6 for cells selecting this region in the fish microenvironment. To better illustrate this a comparison is shown below in Figure 26.

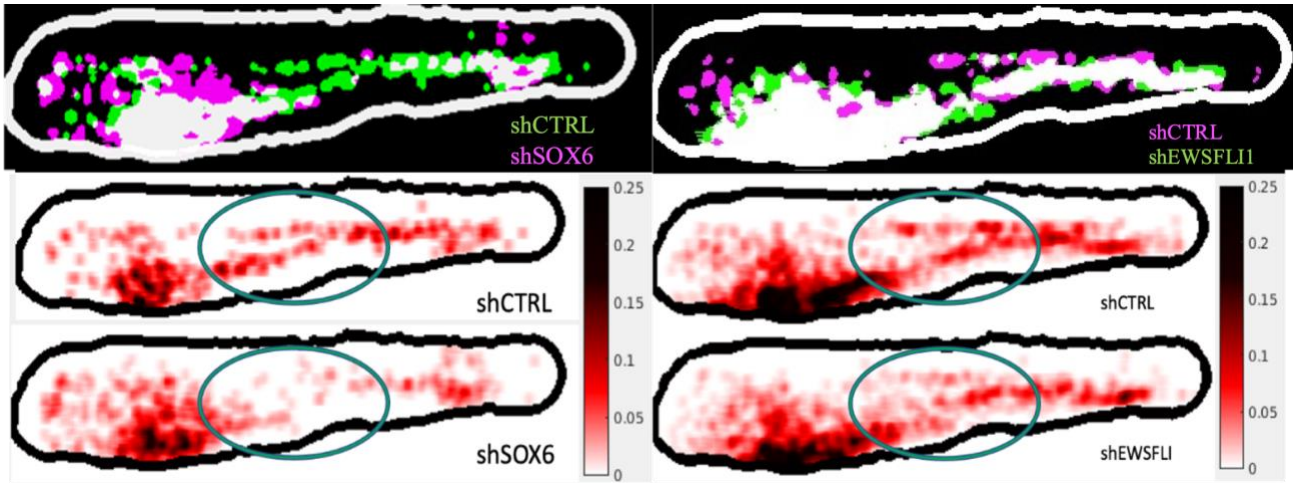



**Figure 27. SOX6 mediated knockdown displays a stronger phenotype by having cell depletion in the central somite space (circle).** SOX6 – N=[54, 96], EWSFLI1- N=[103, 137] for shCTRL and shRNA respectively. Units- Percentage cell residence at a location.

This rather interesting, yet unexpected phenotype is a clear demonstration of the importance of widespread genetic assays. FishATLAS demonstrates a clear interplay between two connected transcription factors in a series of 4 images on the same coordinate template fish. This advancement begets the urgent need to: 1) see how more genes and their targets manifest differences in these metastatic hotspots, 2) follow through on gene-specific drug treatments, and 3) systematically look at shared regions of metastasis across all steps of the mechanism to infer metastatic ‘bastions’ that are consistent across many conditions.

In summary, FishATLAS advances science by showing how and where rare metastasis events, manifest *in vivo*. What remains to be elucidated is a hypothesis that postulates biological mechanisms in EWSFLI1 and its target, SOX6 into making the central somite space and dorsal vein more permissible to metastasis persistence. Moreover, what is

the response of shared, secondary, and persistent areas of metastatic hotspots to drug treatments? These are important questions that require deeper analysis, but currently show a lot of promise for future work. To best analyze these and many other future hypotheses, a rigorous set of statistical tests has been developed for FishATLAS and analysis of global accumulators and local sites in the following chapter.

# CHAPTER FOUR
# Statistical Results of Zebrafish Accumulators

## *4.1 Novel Statistical Testing in Zebrafish Xenografts*

A strong majority of cancer assays have been qualitative measurements of single genetic pathways presenting information in the form of western blots, gel electrophoresis, and others. These have led to great insights, but the need for different tools is paramount for understanding more complex pathologies like metastasis.

FishATLAS seeks to address a key need in the cancer research community of quantifying metastasis in specialized microenvironments *in* vivo through high-throughput, multiple gene screening. This is done using non-parametric statistics, linear assignment optimization problems, and spatial statistics on registered cancer xenograft images. Using image diffeomorphisms as a registration method, image registration of moving objects to template falls on the order of $<20\mu M$, affording confidence for the statistical conclusions on the metastatic microenvironment.

This accuracy translates to the accumulation space where a suite of statistical tests was developed to functionally determine how each condition's interplay works with the previously discovered metastatic hotspots. The suite of statistics available at this current implementation of FishATLAS consists of global cell location permutation testing of both

global and local accumulators, K-Ripley clustering analysis of accumulators, and graph matching/linear assignment of sites between two conditions. Each method tests at progressive levels of granularity the similarities and dissimilarities of tumor microenvironments between biological conditions. To illustrate this, a schematic and key conclusion is given for each part of the statistical test.

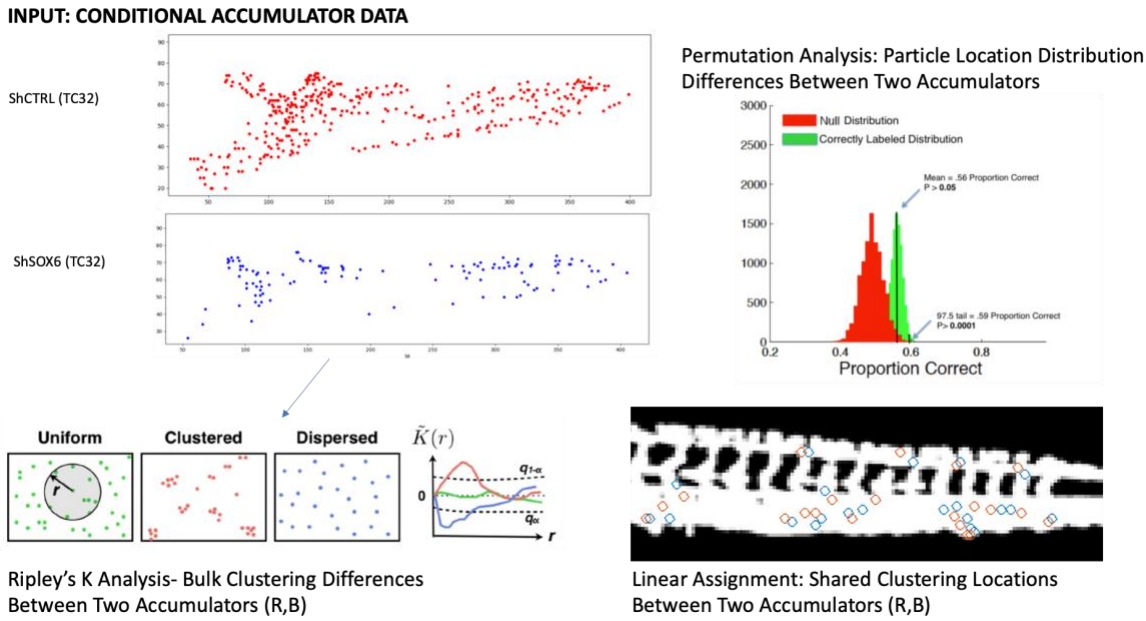



**Figure 28. Schematic overview of the statistical testing methodology for FishATLAS looking at global, local, and matched statistics between genetic conditions.**

Outlined above are simplified outputs from the permutation test, Ripley's K-analysis, and linear assignment of particles. From accumulation, each of the statistical tests above shows a different facet of the many pieces that make up the TME. The upcoming chapters present the data for EWSFLI1 and SOX6 along with melanoma and clonal populations.

## *4.2 Permutation Testing: Global and Localized Significance of Ewing Sarcoma Cells, Clonal Expansions, and Non-Neoplastic Cells*

Permutation testing is used in FishATLAS to determine if two binary masks of cell deposition sites are statistically different across conditions. Permutation testing is agnostic of the number of cell depositions at a given location and is contingent on where each cell is. Although, the simplest of the testing procedures, it is limited in the sense that functionally, the number of cells at a given location would have to be accounted for by setting a cell number threshold. An advantage of using binary mask permutation on a standard template is for a robust definition of local areas based on one template fish's morphology.

### 4.2.1 Determining Differences in Global and Local Microenvironmental Cell Location Distributions

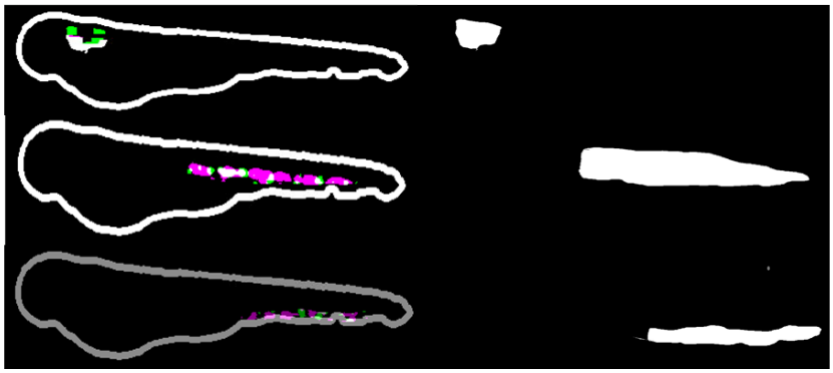


**Figure 29. Standardized microenvironmental masks for localized analysis on a standardized template fish.**

Permutation tests can be calculated on both a global and localized level. Global statistical comparisons are defined as the direct comparison of the whole fish accumulator/s while localized tests utilize a user predefined mask. For example, the following defined local regions are used in FishATLAS for analysis: 1) head, 2) somites, or 3) CHT. Each of the regions and their

respective masks used in the permutation testing workflow is outlined in the adjacent Figure 28.

Both localized region and global permutation tests utilize the same algorithm, allowing for direct comparison to historical data if necessary. Initially, these non-parametric methods work, but future sophisticated methods incorporating user feedback and AI-guided selection are likely to be more effective.

Ideally, for microenvironmental mask delineation, one would be able to use microenvironmental-based segmentation or more direct biological inputs. Current implementations utilize a hand-drawn estimate of each region based on brightfield image morphology. Despite this limitation, these methods are far more advanced than current segmentation methodologies used in mice which cannot ascertain accurate organ location without invasive methods.

### 4.2.2 Global and Local Differences in TC32 Ewing sarcoma and Non-Cancerous Fibroblast Cells

To determine if cells are passively moved by the circulation and not through cancer-unique motility, NIH 3T3 fibroblast cells were injected into 2 dpf zebrafish and imaged one day later. FishATLAS accumulations were completed for fibroblast (N=47) and WT TC32 (N=49) EwS cells. These data are shown below in Figure 29 with global permutation p-value.

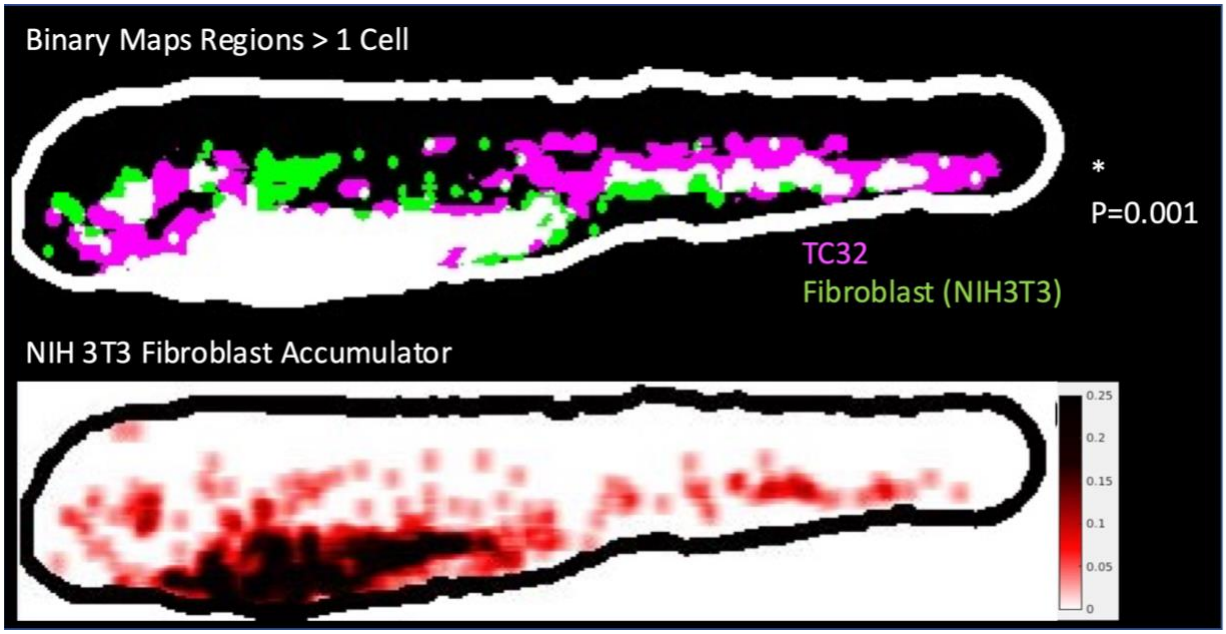



**Figure 30. Non-transformed cells exhibit minimal spreading in FishATLAS accumulation assay.** A) NIH3T3 fibroblast cell accumulator and B) Overlay of fibroblast (green) with TC32 Ewing sarcoma cell (magenta) accumulator.

Imaging accumulation shows a clear visual difference and is supported by permutation tests showing a global statistical significance ($P<0.05$) between cancer cells and fibroblasts. Due to the limited number of fibroblasts present in the CHT and head, localized tests were not completed from the lack of data in those regions. Therefore, the p-value would be a meaningless calculation as there exists no data overlap other than at the injection site.

To summarize, this result is important because it addresses a key question: are metastatic hotspots due to blood circulation or other inherent physics of being in a living organism or is cancer truly unique? The evidence given strongly supports that cancer cells *in vivo* create hotspots that are not caused by physiological *in vivo* motions to cells. Hotspots

are unique to each of the cancers shown, and even display differences in their locations. Such behavior is absent in non-transformed cells as seen in Figure 29.

### 4.2.3 TC32 Clonal Expansions: Global and Local Permutation Tests

To determine how intrinsic heterogeneity of a single clone of EwS could characterize itself in the FishATLAS assay, TC32 clonal expansions were prepared and transduced with a RFP-band reporter. FishATLAS' salient objective of this comparison is to determine if the hotspot distributions seen occur as a product of TC32 nascent heterogeneity, tumor microenvironment, or by the shRNA genetic KD.

Realistically speaking, a conclusion pointing to a complete answer to this question is impossible with the current amount of data. The three-clone data is to initiate this critical question of how metastasis is driven through environment and/or genetics for a tumor heterogeneity. To illustrate the relative similarities in cellular accumulators, a figure of three clones, referred as Clone 7, 9 and 11, are given below along with a global statistical result from permutation testing.

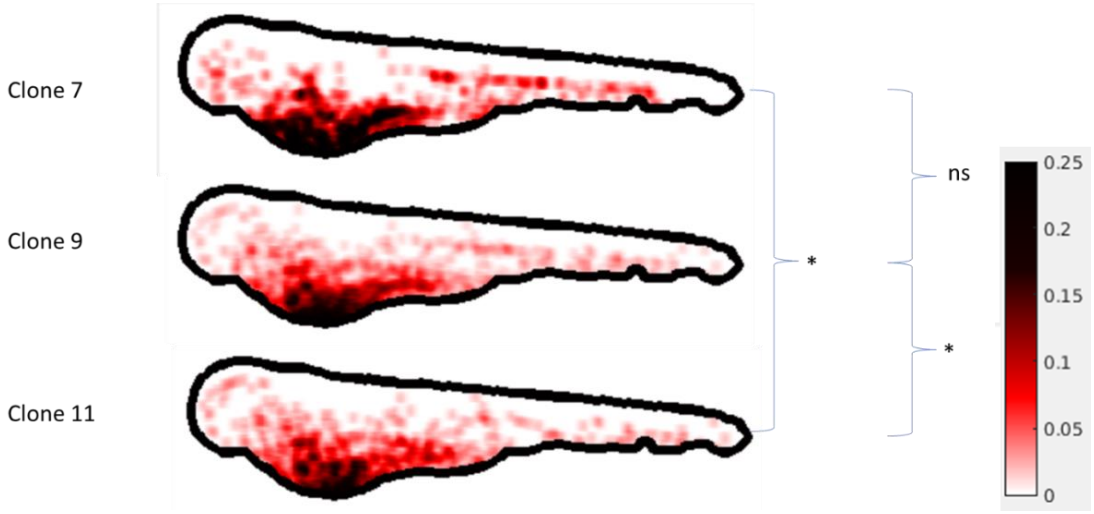



**Figure 31. Clonal expansion accumulators show similarity between 2 out 3 accumulators, but dissimilarity with one. Units- Percentage residence at location**

Global permutation testing shows global similarities between clones 7 and 9, but dissimilarity with clone 11. Globally, this is a mixed result, however particle matching data (presented in succeeding section) between each of the clone accumulators show no statistical differences in the regions where these cells and hotspots accumulate. This was determined by permutation testing of the matched cell-cell distances and tabulating the average matched distance for each clone, with each falling within 1 standard deviation to each other ($p>0.05$, $d_{assignment} = 7.21, 7.4, 7.1$). Fundamentally, these two tests answer different questions about raw location and shared location similarity respectively, so at present it is still inconclusive to determine to a small degree which pressure is favored: cancer heterogeneity and microenvironment. However, these answers will improve with additional clonal experimentation.

In addition, permutation mask testing of each clone was done using the same masking as described with the fibroblast data and is presented below on Table 2.

**Table 2. Pairwise p-value matrix of each clone statistical permutation test shows dissimilarity in Clone 11 to 7 and 9 in the head (top row) somites (middle row) and the CHT (bottom row).**

| *p-Value: Local Permutaiton* | ***Clone 7-9*** | ***Clone 7-11*** | ***Clone 9-11*** |
|---|---|---|---|
| ***Head*** | Undet. | Undet. | Undet. |
| ***Somites*** | 0.089, ns | 0.042, * | 0.02, * |
| ***CHT*** | 0.12, ns | 0.048, * | 0.01,* |

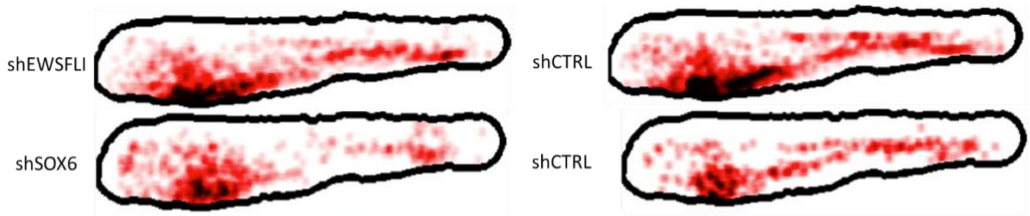
Local mask permutations showed heterogeneity with regards to similar and dissimilar cell deposition locations. First, between clones 7 and 9, there was a high degree of similarity in the global and local accumulators ($p>0.05$). Second, although we can calculate a p-value from a permutation test of the head region, it is not tractable as a sparse event of an already rare event, so the underlying assumptions of giving a large enough sample size would be questionable. Finally, the CHT space showed very little change ($p>=0.05$) between Clones 7 and 9, but differences with 11. Previous data and published work suggest that the CHT is an ideal place to look when doing a comparison of conditions (Tulotta, He et al. 2016) and offer an interesting region for systematic analysis of cancer heterogeneity.

In summary, the statistical data from the clones is the first of its kind to look at interplay between genetic and microenvironmental pressures *in vivo*. Results at raw location show dissimilarity between 1 out of the 3 clones, but all clones show similarity in matched sites. This data isn't enough to conclude anything revolving heterogeneity but gives better understanding for future expansions of clonal accumulators.

### 4.2.4 Statistical Comparison of shEWSFLI1 and shSOX6 accumulators

In the EWSFLI1 and SOX6 accumulation space, each condition was accumulated with a total cohort of 350 fish subdivided into two TC32 shCTRLs and two targeted shRNAs for both EWSFLI1 and SOX6. These conditional accumulators along with the resultant permutation tests are given below on figure 32.

EWSFLI1 and SOX6 conditional accumulators show differences at varying levels between shCTRLs and shRNAs. Both SOX6 and EWSFLI1 KD showed decreased cell




| Permutation Test Region | shEWSFLI vs shCTRL | shSOX6 vs shCTRL |
|---|---|---|
| **Global** | p >> 0.05, ns | p < 0.05, * |
| Head | Inconclusive | Inconclusive |
| Somite | p< 0.05, * | p < 0.05, * |
| CHT | p > 0.05, ns | p < 0.05, * |

**Figure 32. EWSFLI1 displays a localized significance in the somite, however SOX6 displays global significance in all regions except the head. Head region sampling p-values varied greatly, so current values are inconclusive using this injection procedure and method.**

deposition in the central somite space, with a stronger somite removal phenotype seen in the shSOX6 comparison. As a result, it is interesting that the upstream target of SOX6, EWSFLI1 shows little difference between shRNA and shCTRL correspondingly having the lowest significant p-value (p=0.047). Prevailing literature in both *in vitro* and less longitudinal *in vivo* studies show that knockdown of EWSFLI1 not only changes the shape of these cells, but also increases their motility (Segal et. al, 2021). Previous collaborations hypothesize that oxidative stress or cell survival might be playing the dominant role seen in these accumulations (Marchetto, Ohmura et al. 2020). Why these differences manifest in the accumulation space is currently unknown, but multiple lines of investigation are being pursued including oxidative stress and hypoxia.

As a final note, some Gaussian blurring is introduced to the EWSFLI1 accumulation data due to the raw sample size being 40% larger than SOX6. This however doesn't change the statistical assumptions used in the permutation test as each is normalized by its respective sample size.

## *4.3 K-Ripley: Spatial Clustering Activity Across Two Different Accumulators*

Ripley's K is a spatial analysis metric that shows how spatial point patterns relate to each other as a function of increasing radius. Applying this principle to FishATLAS, it is used to typify clustering behavior of an entire conditional accumulator on a single plot. This method shows graphically if cells tend to cluster in groups or have more sparse distributions. These spatial relationships, hereafter referred to as K-plots, allow a direct comparison with either A) randomly assigned cells showing stochastic based clustering or B) between two conditional accumulator's probability distribution. To test cells don't behave as a random particle, a selection of cells for a Ripley's K comparison were sampled in zebrafish vasculature binary masks. An example showing a comparison of random cell points versus a shCTRL from the TC32 cell line is given below on Figure 32.

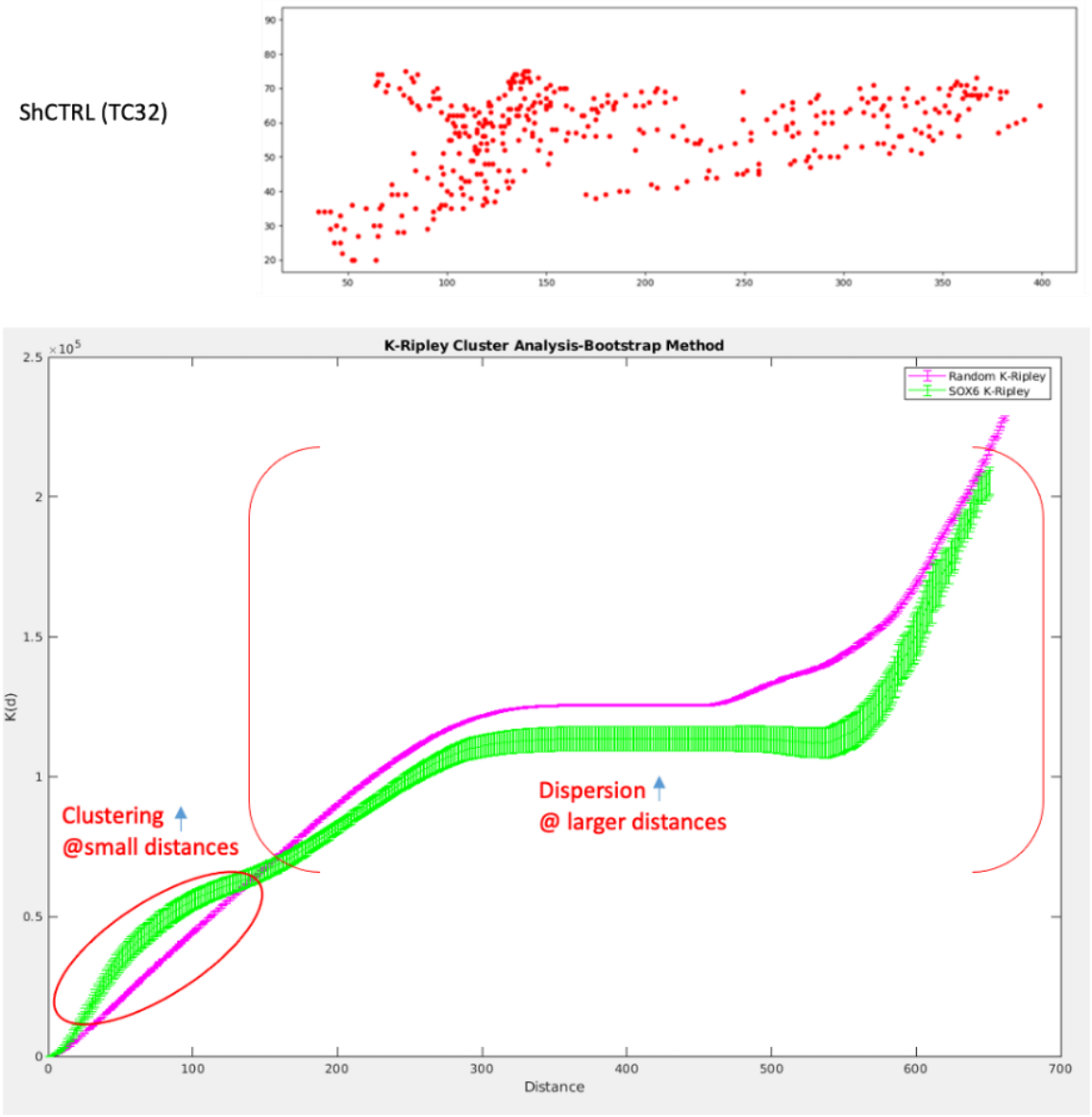



**Figure 32. Clustering of cancer cells is not random in zebrafish vasculature and indicates more clustering at smaller distances in comparison to random.**

It is reassuring that K-plots of global accumulator data show a clear difference between the clustering behavior of the TC32 shCTRL cancer cells and random sampling in zebrafish vasculature. This is seen above on Figure 32 by the increased curvature of the SOX6 (green) line at lower distances above the random (magenta). This served as a

validation step of this assay to demonstrate FishATLAS's accumulations are not random, but grounded in biological mechanisms o. To address this in greater detail, fibroblasts (NIH3T3) were injected into vasculature and plotted in place of the random seeding of cells. This result is shown below on Figure 33.

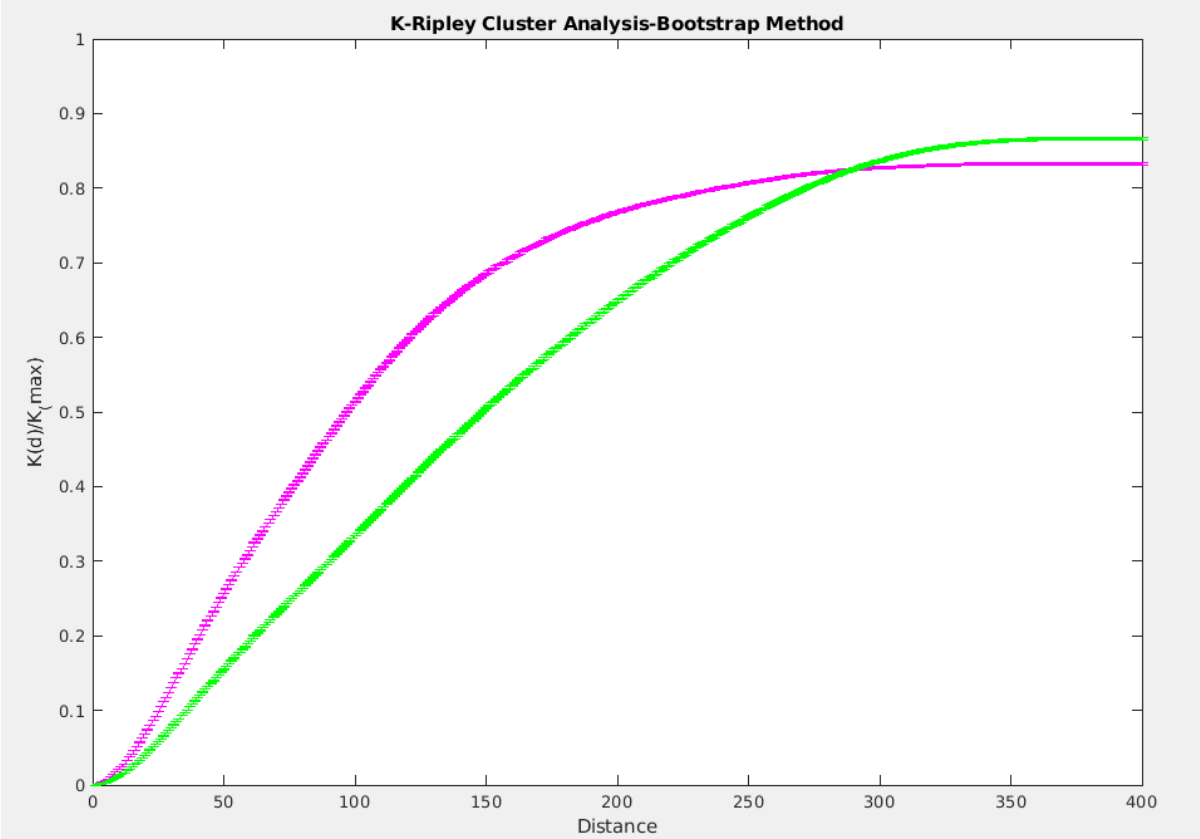



**Figure 33. Fibroblast K-plots show singular larger clustering domain and is different from TC32. NIH 3T3 cells (magenta) K-plot versus TC32 Ewing sarcoma (green).**

Fibroblasts showed different clustering behavior than a random distribution, which was expected because of the physical pressures like circulating blood flow, difference of tissues and other complexities of the *in vivo* environment that are not necessarily encapsulated in random seeding in vasculature.

Repeating this process for shCTRL (TC32 SOX6) as the baseline (maroon) and the experimental group being shSOX6 (green), the following K-plot is obtained on Figure 34:

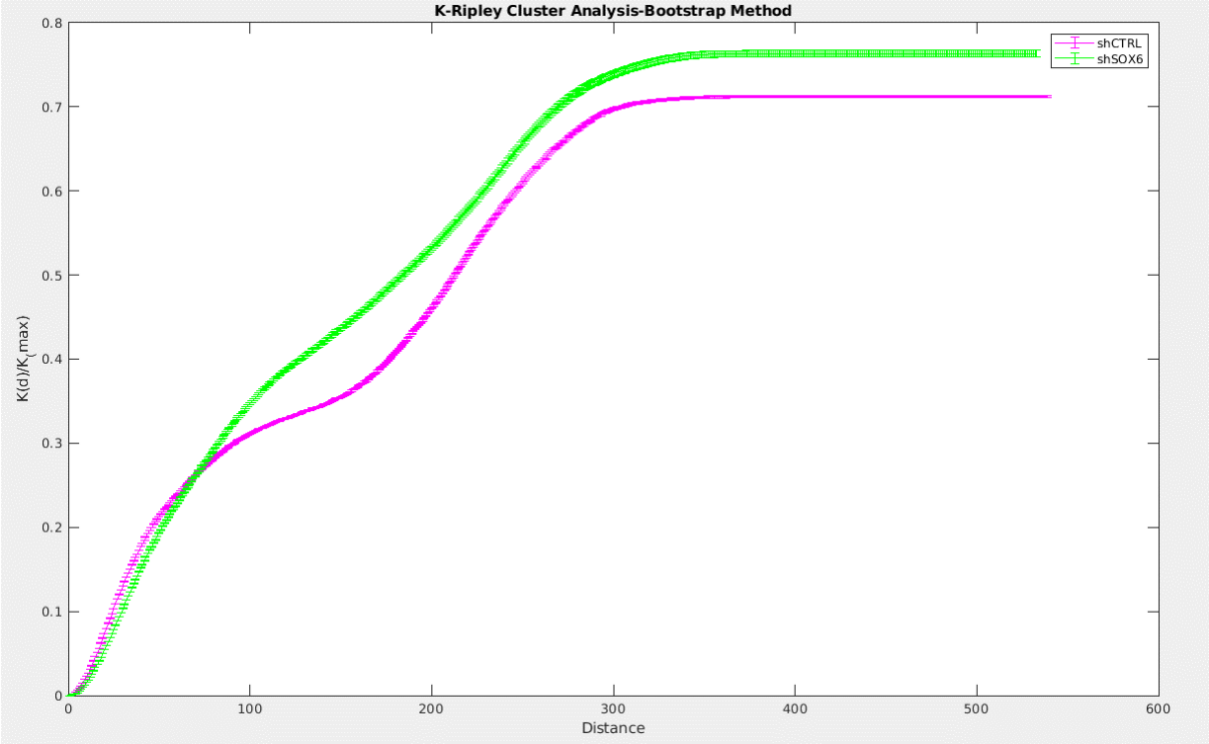



**Figure 34. Ripley's K-function analysis of shCTRL and shSOX6 accumulator show shCTRL data cluster more at smaller distances. TC32 shSOX6 (green) K-curves versus shCTRL (magenta)**

There is a small degree of elevated clustering between cells at lower cell-to-cell distances in shCTRL. This preliminary evidence begins to suggest potential for biologically driven clustering effects that are mediated by SOX6 mechanistically driving higher cell-cell adhesion giving smaller clusters. This could have interplay in many mechanisms with regards to metastatic survival and supports the initial data from simple integrated spreading plots for SOX6 (Chapter 1). This result is both exciting and interesting as it opens many other

pathways of interest for use in this assay. Ultimately, this conclusion would need more data to determine if this is truly the only driver of cell-cell adhesion, which would be difficult and require much more targeted accumulators and data.

## *4.4 Cell Clustering Using Topographical Data Analysis*

Another use of spatial statistics is topographical data analysis persistence homology (TDA). Through Ripley's K function, cell-cell radial expansion distributions are plotted. TDA works differently by showing homological density over different spatial scales and has advantage over Ripley's K function in distinguishing between diverse point patterns (Licon-Salaiz 2020).

TDA is new to the biology field. Most recently, McGuirl et al. used TDA to analyze zebrafish striping patterns to discern phenotypic changes in transgenic lines or response to drug therapies (McGuirl, Volkening et al. 2020). TDA has strides in the field of forestry and geophysics for analyzing spatial statistics of tree distributions, glaciers or other land features (Catania and Neumann 2010). It relies on graphing the  Betti numbers 0 and 1 using a process called persistence homology. Using these Betti numbers, graphically one can discern, on a whole accumulator, a fingerprint of the clustering behavior for comparison to other conditions. Although where it stands as of now has yet to be fully discerned, future iterations of this methodology can use this as another tool for determining metastatic behavior on the accumulation space.

Topographical data analysis has been illustrated on fish, and a schematic is given below on Figure 35 (McGuirl, Volkening et al. 2020).

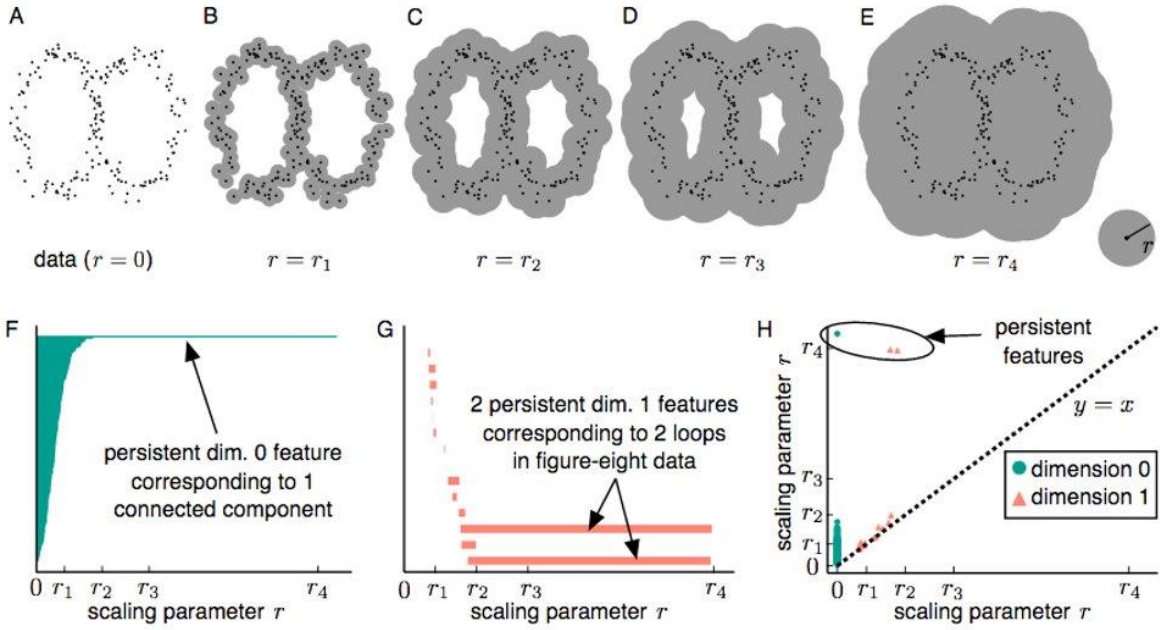



Figure 35**. Increasing radius (A-E) and particle radius tangent events are plotted as a function of increasing radius by number of connected points (F) to processing the persistence homology graph of the Betti numbers 0 (F) and 1 (H).** (McGuirl, Volkening et al. 2020)

The interpretation of the linear (H0) Betti Number might be initially confusing, but it correlates to the amount of radial touches (death) events that occur from all points expanding to r = r' . These results, when graphed give the reader and immediate sense of which distances are most frequented by clusters on a globalized scale that complements the more

finer findings of the Ripley K in addition to taking into account potential shaping issues. Shaping issues for other methods (K function) might come into play for non-pure Cartesian imaging spaces; namely the shape of the fish being relatively complex and not a perfect square. TDA serves as a way to observe similar conclusions as the K-plots without disregarding this assumption, although suspected to be minimal.

Applying TDA to SOX6 accumulator data, the following figure shows the persistence homology for both shSOX6 and shCTRL.

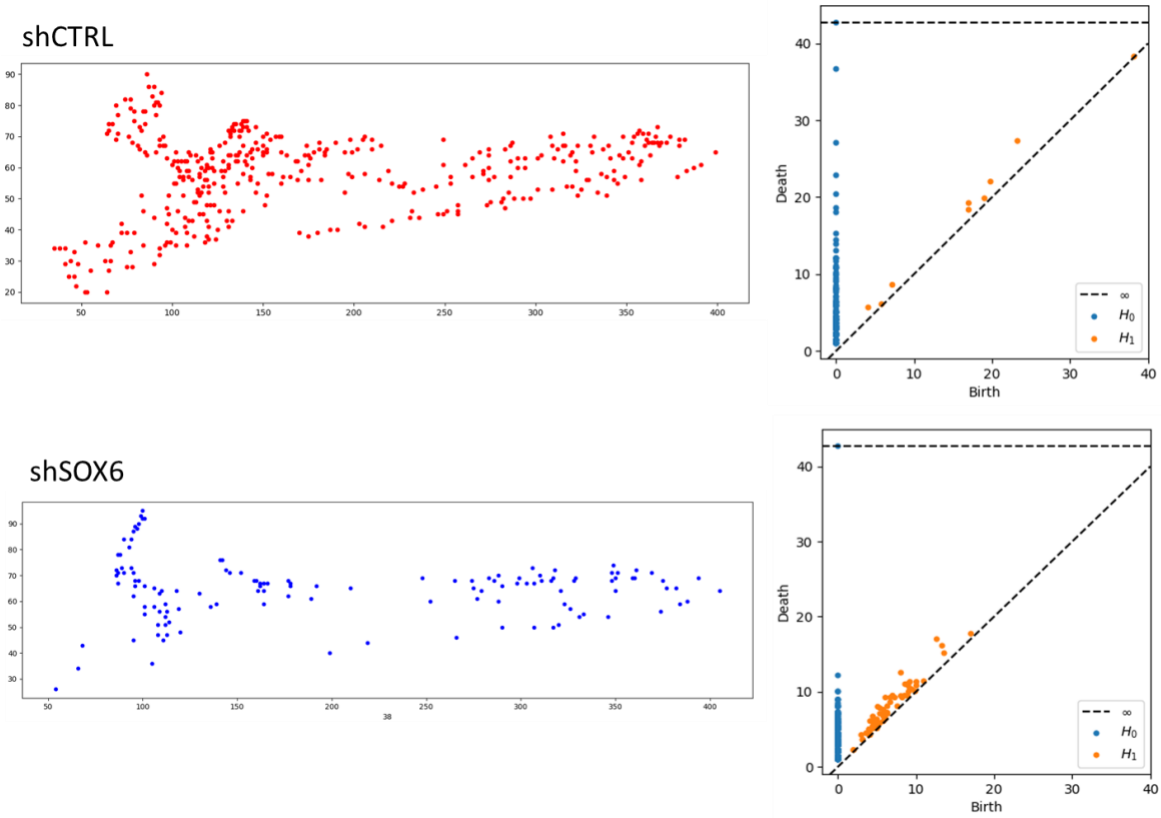



**Figure 36. Persistence homology graphs of SOX6 data show evidence of higher spreading in shCTRL fish at radii ranges R ∈[12, 38]**

Persistence homology results show a clear marked difference between shCTRL and shSOX6 in the first Betti number (H0) through a wider distribution of clustering radius sizes being birthed in shCTRL accumulator from death ranges [12, 40). These death points are absent (12-38 pixels) for shSOX6 H0 graph.

These results in tandem with the K-Ripley results show preliminary evidence for cancer cell cluster removal both at the smaller granularity (localized metastasis testing from TDA) and from a global view (K-Ripley on each cell location.) K-Ripley results show support for a higher clustering presence at lower radii, which appears to be recapitulated through TDA/PH perhaps supporting the hypothesis that SOX6 plays a role in cell survival for these smaller clusters that are removed upon decreased levels of SOX6. Future data and experimentation will show how these effects manifest in additional conditions to bring about some answers to how these clustering effects play into the metastatic cascade and tropism for certain organs/regions.

## *4.5 Linear Assignment: Shared Hotspot Locations Between Accumulators*

Linear assignment (LAP) fundamentally is a matrix linear algebra problem. This process could be thought about as having N workers and then M computers that can be used by these workers. In a case where N>M, there would need to be an optimization of the cost

of not having all workers at a computer, or not all computers having a worker. Evaluation of cost (called the cost matrix) is the key parameter of a linear assignment process and careful selection is required for a given hypothesis. For this project, the cost is distance; FishATLAS wants to determine clusters as close as possible from condition to condition ideally getting in the same TME. Distance is referred to hereafter as the 'cost function.'

Using a cost matrix, the next step is to do an initial filtration of the thousands of cell particle maxima, selecting for clusters that have a minimum of 2 cells in its cluster (as determined through its 2X its peak normalized Gaussian signal value.) These will then create an accumulation of local maxima coordinates that may be used for LAP, however a sense of biological landmarks are important. FishATLAS successfully introduces this landmarking information by registering two compared conditions to a FLI:GFP template zebrafish. This allows for a more realistic portrait of really where these shared sites exist. Taking a typical point distribution below for both shCTRL and shSOX6, these are mapped using an LAP process on Figure 37.

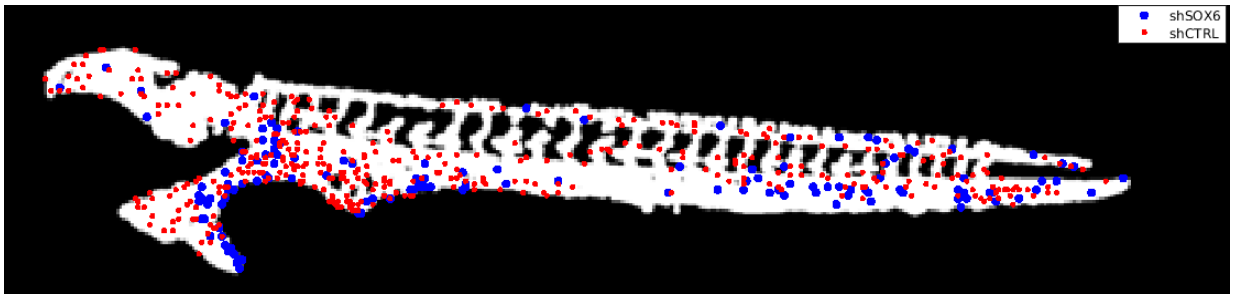



**Figure 37.Cell local maxima distributions show shared (blue) regions of interest in the CHT space. Red= Particle detection event in SOX6 shCTRL or shSOX6 accumulator.**

From a raw particle scatterplot in zebrafish data, regions like the CHT and injection space show a high degree of overlap of cell local maxima. What remains unclear is if this is the total list of shared sites and what regions at a finer level share a metastatic hotspot.

LAP was initiated using a distance cost matrix that set all distances greater than 10 pixels away as too costly and matched each hotspot with its nearest neighbor in the other condition. This is illustrated below on Figure 38 with shared region (blue) dots showing its effective radius and all detection dots (red).

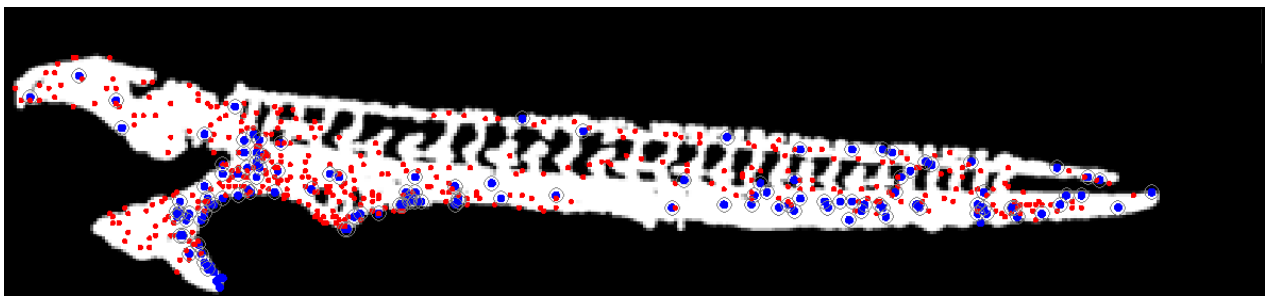


**Figure 38. Nearest neighbor searching of shSOX6 and shCTRL show additional sights of interest in the head, dorsal vasculature, and CHT. Blue spots- Shared site between**

**conditions with nearest neighbor searching radius, Red-All local maxima between conditions)**

Predictably, the regions that were easiest to spot by looking at the raw data, such as the injection space and CHT , show a large amount of shared regions of deposition. Less frequent spaces like the dorsal vasculature and tail inter-somitic arteries were less apparent and constitute an interesting TME for analysis. Although this is still relatively new as an analysis, gainful information showing these inter-somitic shared sites suggest this to be a site of higher or altered oxidative stress (Mourabit, Fitzgerald et al. 2019). This presents a novel finding outside of the established interesting TME of the CHT, which has been well established in literature in zebrafish xenografts due to its hematopoietic maturation and signaling pathways active at the time of xenograft (Li, Zou et al. 2014, Xue, Lv et al. 2017) .

### 4.5.1 Distance Metrics Between Linear Assignment Matching

A similar application of LAP is through analysis of the average assignment distance. To make this a meaningful quantity for comparison, particle assignments from two different conditions are matched, and their corresponding cell-cell distances averaged to generate a single scalar value for permutation testing. These may be used to draw a comparison of locations of shared clustering locations (Wei, Alsaadi et al. 2017) . Below on Table 3, is a summary of linear assignment distance statistical testing for Ewing sarcoma along with corresponding accumulator images (Figure 39).

**Table 3. Integrated pair-wise cell distances from linear assignment matching for TC32 cells under different biological accumulations. Color scale represents close (green), moderate (yellow), and long (red, scarlet) average cell distance. (shCTRL-EWSFLI1 control and shCTRL_2-SOX6 scrambled control, *indicates statistically significant difference)**

| ***Dist.(pixels)*** | **shCTRL** | **Clone 7** | **Clone 9** | **Clone 11** | **shCTRL_2** | **shSOX6** |
|---|---|---|---|---|---|---|
| *shEWSFLI1* | 5.64 | 26.84* | 24.94* | 22.70* | 27.43* | 27.24* |
| *shCTRL* | X | 23.43* | 21.72* | 18.53* | 25.69* | 24.93* |
| *Clone 7* | X | X | 7.44 | 7.71 | 24.45* | 15.64* |
| *Clone 9* | X | X | X | 7.84 | 21.13* | 15.64* |
| *Clone 11* | X | X | X | X | 21.14* | 13.04* |
| *shCTRL_2* | X | X | X | X | X | 18.36* |

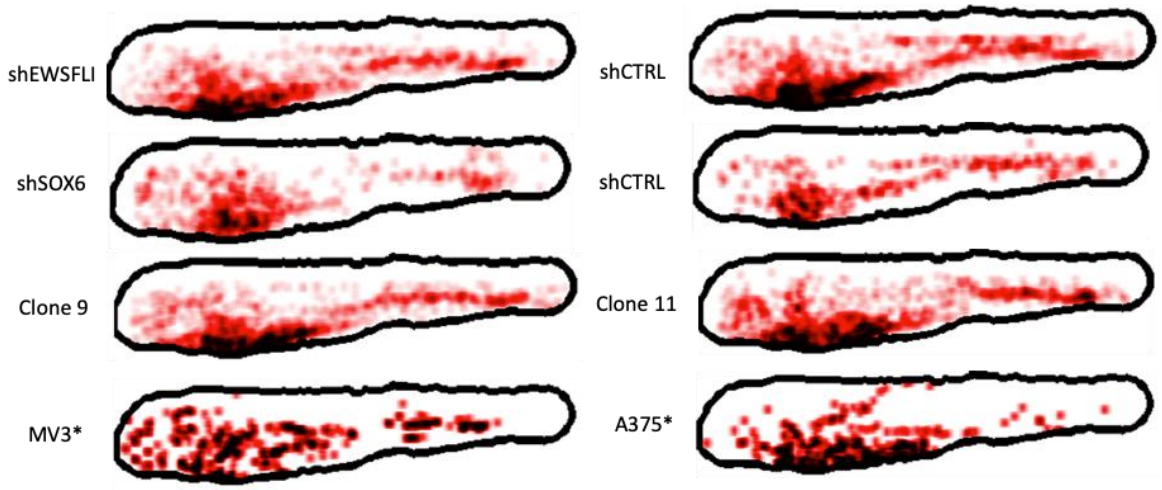



**Figure 39. Accumulation conditions across all different conditions illustrating the heterogeneity of hotspots and differences between cell of origin and genetic condition.**

This preliminary result offers interesting and complicating conclusions. For one, it is surprising that the overall distance between matches, is statically unsignificant between shEWSFLI1 and shCTRL (EWSFLI control). Both appear visually similar except for the dorsal vasculature, lending to suggest that for some conditions a more sensitive statistic might be required.

Between clonal expansions, as mentioned in prior sections, the average distance was unsignificant based on permutations suggesting that shared sites between all 3 clones are similar. How this translates in addition to the global permutations being similar with 2 out 3 accumulations is unclear, but will likely get elucidated once future clones are generated.

SOX6 data of shSOX6 and its shCTRL show further evidence of a significant difference in both global permutations and shared sites of metastasis. What will have to be demonstrated in future data is how this will change with more sensitive matched site analysis pipelines and with changes in drug treatment. In addition, it presents the very perplexing question of why. Why is a downstream target of EWSFLI1 having such a strong effect on metastatic hotspots, their shared/secondary sites, and potentially microenvironment. Although evidence of a strong difference has been obtain, many more questions have been opened that could begin to determine how metastatic properties are changed by this target and its microenvironment

## *4.6 Concluding Statements on FishATLAS Statistical Methods*

In summary, FishATLAS statistics focus on three different levels of granularity:

1. Location Masks- Global and localized microenvironments of interest
2. Global Clustering Behavior- Determining at a global level if cell locations are clustered at a given distance
    a. Ripley's K Analysis- Global clustering events at all radial distances in comparison to random sampling and other conditional accumulators
    b. Topographical Data Analysis (TDA) – Using persistence homology to determine clustering and ring structure fingerprinting as an additional metric to analyze clustering phenotypes.
3. Shared Cell of Metastatic Hotspots-Determining shared sites and microenvironments using Linear Assignment Nearest Neighbor searching in addition to vasculature landmarks
    a. Graph/bipartite matching of cells from two conditions looking at average distance between cell matches.

FishATLAS poses itself in a great position as a software to aid longitudinal studies of a myriad of cancers to functionally interrogate certain suspect genes and their roles in the metastatic cascade. Although these preliminary data show promise for EwS for use in later drug studies, these methods can be used for any cancer or pathology to analyze *in vivo* data.

Statistically, FishATLAS has shown that A) metastatic hotspots exist and distributions are statistically different across different conditions locally and globally B) EwS displays different cell clustering behaviors upon KD of SOX6 and finally C) Shared metastatic sites exist in the CHT and inter-somitic arteries in both SOX6 KD/WT in addition to other conditions.

Many interesting questions remain regarding how these accumulation images and statistical tests change with response to drug therapy. In addition, the addition of biosensors for both the cells and fish microenvironment have yet to be fully exploited beyond that of fluorescent reporters. To truly begin to understand how genetics and microenvironment come to drive organotropism, these are necessary areas for future research.

# CHAPTER FIVE
# Conclusions and Recommendations

## 5.1 FishATLAS: A pipeline for genetic and environmental visualization *in vivo*

As previously mentioned, the interplay of both environment and genes mirrors a central question of nature or nurture being a driving force, and it does appear that environment plays a large part. FishATLAS both addresses the aims of the project and has generated many more hypotheses as to the interplay of cells and their genetic and environmental pressures. These aims and their progress are given below.

1) **Determine the existence of metastatic hotspots for human zebrafish xenografts**
    a) **Functionally determine differences in hotspot deposition between EwS, melanoma, non-transformed cells, and EwS clonal expansions**

Through development of the diffeomorphic registration and novel particle detection pipeline, we can systematically **determine the existence and differential presence of hotspots *in vivo*.**

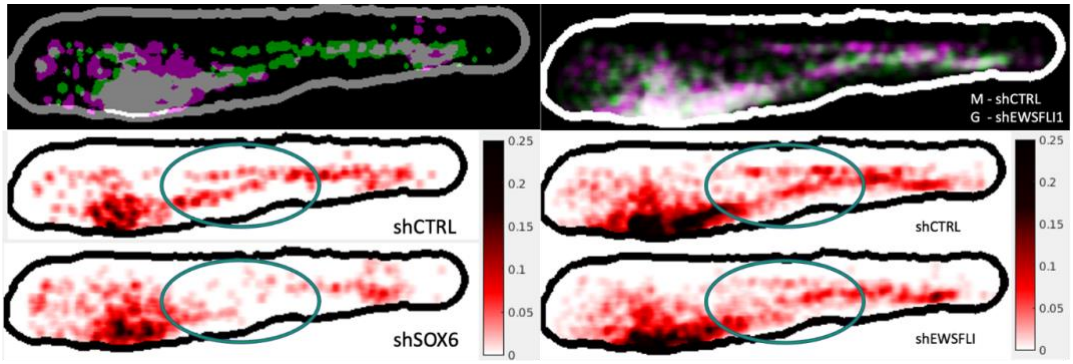



**Figure 40. Aim 1: FishATLAS demonstrates the existence of metastatic hotspots and their different distributions between different genetic pressures.**

The research findings for the presence of hotspots were more than expected. To see clear hotspots of 20% frequency in large data cohorts is a feat that is both novel, and impactful for the field of cancer biology. We can directly image areas the interplay of genetics and microenvironment on a single image, and by utilizing statistics, begin to functionally probe the very complex problem of metastasis and cancer. This is been shown in multiple different types of cancer as well in Figure 41.

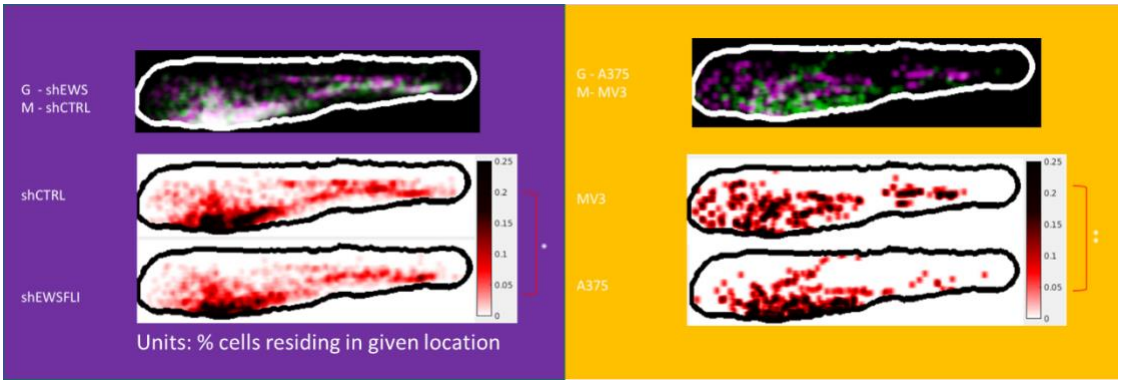



**Figure 41.Different cancer accumulations display differences in hotspot location. Purple-TC32 Ewing sarcoma shEWSFLI1, Yellow- MV3 (top) A375 (bottom) melanoma datasets.**

With the discovery of hotspots, the next goal was to determine the other drivers behind these distributions. To start, how does the intrinsic heterogeneity of the tumor effect the distribution of the hotspots? This was the reasoning for Aim 2 and is discussed below.

**2) Initiate investigating the role between intrinsic heterogeneity and microenvironmental pressure.**

**a) Experiment on TC32 EwS monoclonal expansions to initiate investigations on clonal effects and environmental pressure**

To begin to ascertain whether the intrinsic heterogeneity is the cause for these hotspot phenotypes, a series of 3 clonal expansions was completed. The goal of these was to determine the overall feasibility of FishATLAS in seeing similarities and differences in the same clone. These data are given below:

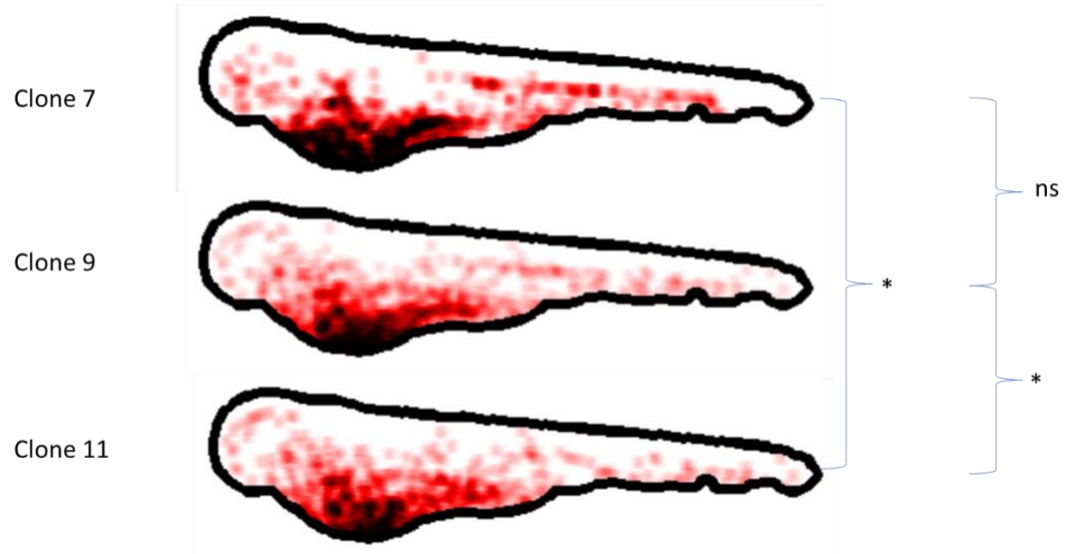



**Figure 42. TC32 clonal expansion accumulators show global similarity between clones 7 and 9, but significant difference ($p<0.05$) with Clone 11. Local pairwise matching of clone show similarity between all clonal data ($p>0.05$).**

Clonal expansions results were mixed but demonstrated need for expansion of clonal data to probe the dependency of metastasis on intrinsic heterogeneity. Despite similarity being found mixed at global and clones being similar in local testing, this hypothesis needs much more data to determine how tumor heterogeneity of the cancer is more dominant over environment or if the converse is true.

**3) Determine conserved areas of metastasis between conditions to serve as targets of interest in cancer therapeutic studies.**

a) Functionally map conserved areas onto biological landmarks

Accumulation spaces are very useful for determining global and local trends to an extent, but where do singular events fall *in vivo* in morphology. To illustrate this point, accumulated images displaying hotspots in areas of interest were registered back to fluorescent vasculature reporter templates in the zebrafish. This is shown below:

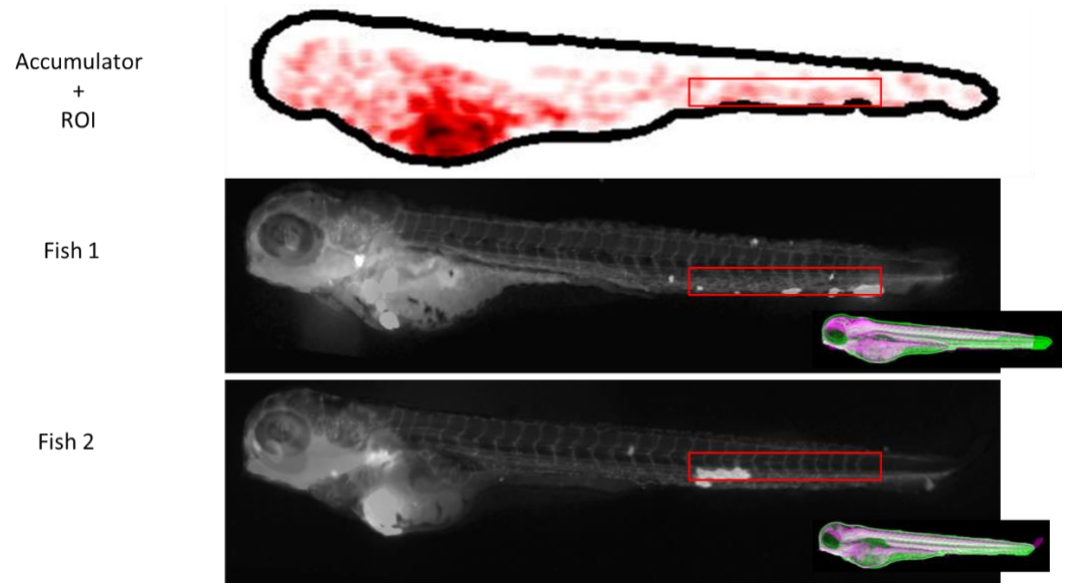



**Figure 43. Environmental mapping of outlined hotspot ROI on zebrafish show single fish data of deposition sites to inform specific microenvironmental data.**

This astounding result shows that the diffeomorphic registration and detection pipeline is quite adept at single fish. This is particularly important because of the potential for determining secondary sites of metastasis or using other transgenic labels to delineate organ boundaries or any other biosensor. This is a large scaling point for FishATLAS' plans in determining secondary sites of metastasis.

Although not fully elucidated yet, there is promising work currently being done for this project in using pair-wise distances to determine locally if two conditional accumulators have shared or different metastatic hotspot distributions. The results of these are initially promising and are summarized below on Table 4.

**Table 4. Integrated pair-wise cell distances from linear assignment matching for TC32 cells under different biological accumulations. Color scale represents close (green), moderate (yellow), and long (red, scarlet) average cell distance. (shCTRL-EWSFLI1 control and shCTRL_2-SOX6 scrambled control, *indicates statistically significant difference)**

| ***Dist.(pixels)*** | **shCTRL** | **Clone 7** | **Clone 9** | **Clone 11** | **shCTRL_2** | **shSOX6** |
|---|---|---|---|---|---|---|
| *shEWSFLI1* | 5.64 | 26.84* | 24.94* | 22.70* | 27.43* | 27.24* |
| *shCTRL* | X | 23.43* | 21.72* | 18.53* | 25.69* | 24.93* |
| *Clone 7* | X | X | 7.44 | 7.71 | 24.45* | 15.64* |
| *Clone 9* | X | X | X | 7.84 | 21.13* | 15.64* |
| *Clone 11* | X | X | X | X | 21.14* | 13.04* |
| *shCTRL_2* | X | X | X | X | X | 18.36* |

To help make better sense of these results, a figure with each conditions accumulator is given below:

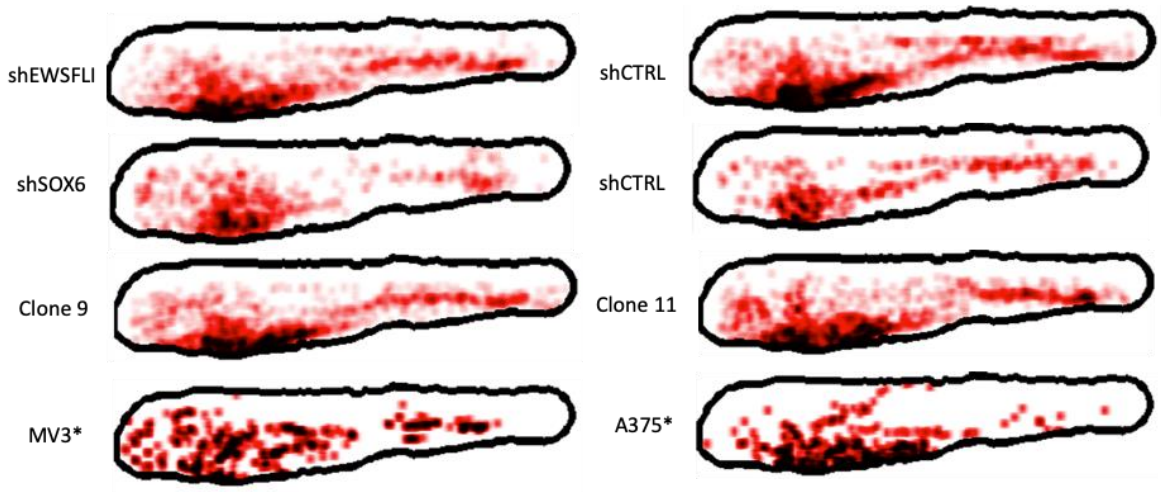



**Figure 44. Summary of all conditional accumulators in FishATLAS assay demonstrating heterogeneity of hotspot location.**

These pair wise distances appear to be quite a powerful tool to determine if there are shared sites of metastasis. The results surrounding the EWSFLI1 data appear to be non-significant locally, which is a departure from literature (Franzetti, Laud-Duval et al. 2017). However, the importance of this conclusion states the need of a full analysis of the used TC32 cell line via genotyping or potentially use another cell line with the same construct to confirm the result. Overall, the results constitute great progress in the field for a pipeline that looks at the genetic and microenvironmental pressures to begin to understand the organotrophic behaviors of cancer.

In summary, FishATLAS constitutes a huge step forward in cancer research for patients with metastatic disease. Cancer itself proves itself to be a cacophony of many genes and extrinsic signals  that require both focused and widespread genetic screening as well as microenvironmental information.

What this project has outlined was just how important and different each gene's perturbation is to metastatic deposition and that such a complex process like metastasis cannot be defined by a single gene or microenvironmental stimulus. These processes need to be evaluated systematically and require data statistics to analyze each of these sparse data events. To do this feasibly requires use of the zebrafish system along with the methodologies described here.

It is the sincerest hope that FishATLAS paves the way to future research for many labs and researchers to begin to determine enviro-genetic mechanisms that inform treatment decisions for patients and people.

# BIBLIOGRAPHY

Borah, B. and D. K. Bhattacharyya (2004). An improved sampling-based DBSCAN for large spatial databases. International conference on intelligent sensing and information processing, 2004. proceedings of, IEEE.

Catania, G. A. and T. Neumann (2010). "Persistent englacial drainage features in the Greenland Ice Sheet." Geophysical Research Letters **37**(2).

Chefd'Hotel, C., et al. (2002). Flows of diffeomorphisms for multimodal image registration. Proceedings IEEE International Symposium on Biomedical Imaging, IEEE.

Criscuolo, M., et al. (2020). "Family functioning, coparenting, and parents' ability to manage conflict in adolescent anorexia nervosa subtypes." Fam Syst Health **38**(2): 151-161.

INTRODUCTION: Studies about family functioning in adolescent anorexia nervosa (AN) are conflicting and often do not consider the possible differences between the restricting and binge/purging subtypes (AN-R and AN-B/P). Moreover, an underestimated element is the quality of the coparenting relationship, that is, the level of coordination and agreement between parents, as well as the methods of managing parental conflict. METHOD: The study aims to explore family functioning, coparenting and conflict management in a sample of 60 adolescents with AN-R and AN-B/P and their relations with AN severity in terms of body mass index (BMI). Patients and parents completed the following questionnaires: Family Adaptability and Cohesion Evaluation Scales, Coparenting Scale-Revised, and Conflict Management Questionnaire. RESULTS: No differences are found in general family functioning between the two groups, but mothers in the AN-R group show a higher coparental conflict and a tendency to use negative-passive ways of conflict management than mothers of the AN-B/P group. Moreover, the presence of a coparental conflict is associated to lower BMI and to more dysfunctional family functioning. DISCUSSION: These findings underline the importance of increasing knowledge on specific aspects of general family functioning. In particular, it seems that the way parents manage conflict between themselves as partners can have a role in the maintenance of AN. Clinical implications of these results are discussed. (PsycInfo Database Record (c) 2020 APA, all rights reserved).

Deriche, R. (1990). "Fast algorithms for low-level vision." IEEE transactions on pattern analysis and machine intelligence **12**(1): 78-87.

Fazio, M., et al. (2020). "Zebrafish patient avatars in cancer biology and precision cancer therapy." Nat Rev Cancer **20**(5): 263-273.

Franzetti, G. A., et al. (2017). "Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells." Oncogene **36**(25): 3505-3514.

Gaudenzi, G., et al. (2017). "Patient-derived xenograft in zebrafish embryos: a new platform for translational research in neuroendocrine tumors." Endocrine **57**(2): 214-219.

Grunewald, T. G. P., et al. (2018). "Ewing sarcoma." Nat Rev Dis Primers **4**(1): 5.

He, S., et al. (2012). "Neutrophil-mediated experimental metastasis is enhanced by VEGFR inhibition in a zebrafish xenograft model." The Journal of pathology **227**(4): 431-445.

He, S., et al. (2012). "Neutrophil-mediated experimental metastasis is enhanced by VEGFR inhibition in a zebrafish xenograft model." The Journal of Pathology **227**(4): 431-445.

Abstract Inhibition of VEGF signalling effectively suppresses localized tumour growth but accelerates tumour invasiveness and micrometastasis by unknown mechanisms. To study the dynamic and reciprocal interactions between tumour cells and their microenvironment during these processes, we established a xenograft model by injecting tumour cells into the blood circulation of transparent zebrafish embryos. This reproducibly results in rapid simultaneous formation of a localized tumour and experimental micrometastasis, allowing time?resolved imaging of both processes at single?cell resolution within 1 week. The tumour vasculature was initiated de novo by remodelling of primitive endothelial cells into a functional network. Roles of myeloid cells in critical tumourigenesis steps such as vascularization and invasion were revealed by genetic and pharmaceutical approaches. We discovered that the physiological migration of neutrophils controlled tumour invasion by conditioning the collagen matrix and forming the metastatic niche, as detected by two?photon confocal microscopy and second harmonic generation. Administration of VEGFR inhibitors blocked tumour vascularization and a localized tumour growth but enhanced migration of neutrophils, which in turn promoted tumour invasion and formation of micrometastasis. This demonstrates the in vivo cooperation between VEGF signalling and myeloid cells in metastasis and provides a new mechanism underlying the recent findings that VEGFR targeting can promote tumour invasiveness. Copyright ? 2012 Pathological Society of Great Britain and Ireland. Published by John Wiley & Sons, Ltd.

Heilmann, S., et al. (2015). "A quantitative system for studying metastasis using transparent zebrafish." Cancer research **75**(20): 4272-4282.

Metastasis is the defining feature of advanced malignancy, yet remains challenging to study in laboratory environments. Here we describe a high-throughput zebrafish system for comprehensive, in vivo assessment of metastatic

biology. First, we generated several stable cell lines from melanomas of transgenic mitfa-BRAF(V600E);p53(−/−) fish. We then transplanted the melanoma cells into the transparent casper strain to enable highly quantitative measurement of the metastatic process at single cell resolution. Using computational image analysis of the resulting metastases, we generated a metastasis score, μ, that can be applied to quantitative comparison of metastatic capacity between experimental conditions. Furthermore, image analysis also provided estimates of the frequency of metastasis-initiating cells (~1/120,000 cells). Finally, we determined that the degree of pigmentation is a key feature defining cells with metastatic capability. The small size and rapid generation of progeny combined with superior imaging tools make zebrafish ideal for unbiased high-throughput investigations of cell-intrinsic or microenvironmental modifiers of metastasis. The approaches described here are readily applicable to other tumor types and thus serve to complement studies also employing murine and human cell culture systems.

Heilmann, S., et al. (2015). "A Quantitative System for Studying Metastasis Using Transparent Zebrafish." Cancer Res **75**(20): 4272-4282.

Heilmann, S., et al. (2015). "A quantitative system for studying metastasis using transparent zebrafish." Cancer research **75**(20): 4272-4282.

Hill, D., et al. (2018). "Embryonic zebrafish xenograft assay of human cancer metastasis." F1000Res **7**: 1682.

Kim, A. and M. S. Cohen (2016). "The discovery of vemurafenib for the treatment of BRAF-mutated metastatic melanoma." Expert opinion on drug discovery **11**(9): 907-916.

Kohlberg, L. (1978). "Revisions in the theory and practice of moral development." New Directions for Child and Adolescent Development **1978**(2): 83-87.

Abstract A questioning of old beliefs that moral stages can provide the sole basis for moral education and that moral indoctrination is necessarily invalid, as well as new views of the moral role of the community and the status of stage 6.

Konantz, M., et al. (2012). "Zebrafish xenografts as a tool for in vivo studies on human cancer." Annals of the New York Academy of Sciences **1266**(1): 124-137.

Lee, L. M., et al. (2005). "The fate of human malignant melanoma cells transplanted into zebrafish embryos: assessment of migration and cell division in the absence of tumor formation." Developmental dynamics: an official publication of the American Association of Anatomists **233**(4): 1560-1570.

Li, F., et al. (2014). "CCL21/CCR7 axis activating chemotaxis accompanied with epithelial-mesenchymal transition in human breast carcinoma." Med Oncol **31**(9): 180.

Secondary lymphoid tissue chemokine (SLC/CCL21) and its receptor CCR7 have been implicated in lymph node metastasis, whereas the mechanism of which remains unclear. Epithelial-mesenchymal transition (EMT) plays an important role in invasion and migration of cancer cells. We presumed that CCL21/CCR7 axis activates EMT process to induce cancer cell invasion and metastasis. Firstly, the expressions of CCR7 and EMT markers were examined by immunohistochemical staining in the primary breast carcinoma tissues from 60 patients who underwent radical mastectomy. Then, we investigated whether CCL21/CCR7 induces EMT process during mediating cancer cell invasion or migration in vitro. By immunohistolochemistry, high expressions of CCR7, Slug and N-cadherin were seen in 60, 65, and 76.67 % of tumors, respectively, and significantly associated with lymph node metastases as well as clinical pathological stage. Furthermore, the CCR7 expression was significantly correlated to Slug and N-cadherin. In vitro, stimulating breast cancer cell lines 1428, MCF-7 and MDA-MB-231 with CCL21, the invasion and migration of tumor cells were promoted, and simultaneously, EMT phenotype of tumor cells was enhanced, including down-regulation of E-cadherin, up-regulation of Slug, Vimentin and N-cadherin at both protein and mRNA levels. Inversely, knockdown of CCR7 by shRNA suppressed tumor cell invasion, migration and EMT phenotype induced by CCL21. These results indicated that CCL21/CCR7 axis could activate EMT process during chemotaxis of breast carcinoma cells.

Licon-Salaiz, J. (2020). Topological Aspects of Turbulence in the Planetary Boundary Layer, Universität zu Köln.

Marchetto, A., et al. (2020). "Oncogenic hijacking of a developmental transcription factor evokes vulnerability toward oxidative stress in Ewing sarcoma." Nat Commun **11**(1): 2423.

Ewing sarcoma (EwS) is an aggressive childhood cancer likely originating from mesenchymal stem cells or osteo-chondrogenic progenitors. It is characterized by fusion oncoproteins involving EWSR1 and variable members of the ETS-family of transcription factors (in 85% FLI1). EWSR1-FLI1 can induce target genes by using GGAA-microsatellites as enhancers.Here, we show that EWSR1-FLI1 hijacks the developmental transcription factor SOX6 - a physiological driver of proliferation of osteo-chondrogenic progenitors - by binding to an intronic GGAA-microsatellite, which promotes EwS growth in vitro and in vivo. Through integration of transcriptome-profiling, published drug-screening data, and functional in vitro and in vivo experiments including 3D and PDX models, we discover that constitutively high SOX6 expression promotes elevated levels of oxidative stress that create a therapeutic vulnerability toward the oxidative stress-inducing drug Elesclomol.Collectively, our results exemplify how aberrant activation of a developmental transcription factor by a dominant oncogene can promote malignancy, but provide opportunities for targeted therapy.

McGuirl, M. R., et al. (2020). "Topological data analysis of zebrafish patterns." Proc Natl Acad Sci U S A **117**(10): 5113-5124.

Mohan, A. S., et al. (2019). "Enhanced Dendritic Actin Network Formation in Extended Lamellipodia Drives Proliferation in Growth-Challenged Rac1(P29S) Melanoma Cells." Dev Cell **49**(3): 444-460 e449.

Mourabit, S., et al. (2019). "New insights into organ-specific oxidative stress mechanisms using a novel biosensor zebrafish." Environ Int **133**(Pt A): 105138.

BACKGROUND: Reactive oxygen species (ROS) arise as a result from, and are essential in, numerous cellular processes. ROS, however, are highly reactive and if left unneutralised by endogenous antioxidant systems, can result in extensive cellular damage and/or pathogenesis. In addition, exposure to a wide range of environmental stressors can also result in surplus ROS production leading to oxidative stress (OS) and downstream tissue toxicity. OBJECTIVES: Our aim was to produce a stable transgenic zebrafish line, unrestricted by tissue-specific gene regulation, which was capable of providing a whole organismal, real-time read-out of tissue-specific OS following exposure to a wide range of OS-inducing environmental contaminants and conditions. This model could, therefore, serve as a sensitive and specific mechanistic in vivo biomarker for all environmental conditions that result in OS. METHODS: To achieve this aim, we exploited the pivotal role of the electrophile response element (EpRE) as a globally-acting master regulator of the cellular response to OS. To test tissue specificity and quantitative capacity, we selected a range of chemical contaminants known to induce OS in specific organs or tissues, and assessed dose-responsiveness in each using microscopic measures of mCherry fluorescence intensity. RESULTS: We produced the first stable transgenic zebrafish line Tg (3EpRE:hsp70:mCherry) with high sensitivity for the detection of cellular RedOx imbalances, in vivo in near-real time. We applied this new model to quantify OS after exposure to a range of environmental conditions with high resolution and provided quantification both of compound- and tissue-specific ROS-induced toxicity. DISCUSSION: Our model has an extremely diverse range of potential applications not only for biomonitoring of toxicants in aqueous environments, but also in biomedicine for identifying ROS-mediated mechanisms involved in the progression of a number of important human diseases, including cancer.

Sha, D., et al. (2020). "Tumor mutational burden as a predictive biomarker in solid tumors." Cancer discovery **10**(12): 1808-1825.

Tulotta, C., et al. (2016). Imaging of Human Cancer Cell Proliferation, Invasion, and Micrometastasis in a Zebrafish Xenogeneic Engraftment Model. Zebrafish: Methods and Protocols. K. Kawakami, E. E. Patton and M. Orger. New York, NY, Springer New York**:** 155-169.

The xenograft model, using the early life stages of the zebrafish, allows imaging of tumor cell behavior both on a single cell and whole organism level, over time, within a week. This robust and reproducible assay can be used as an intermediate step between in vitro techniques and the expensive, and time consuming, murine models of cancer invasion and metastasis.

Tulotta, C., et al. (2016). "Imaging Cancer Angiogenesis and Metastasis in a Zebrafish Embryo Model." Adv Exp Med Biol **916**: 239-263.

Tumor angiogenesis and metastasis are key steps of cancer progression. In vitro and animal model studies have contributed to partially elucidating the mechanisms involved in these processes and in developing therapies. Besides the improvements in fundamental research and the optimization of therapeutic regimes, cancer still remains a major health threatening condition and therefore the development of new models is needed. The zebrafish is a powerful tool to study tumor angiogenesis and metastasis, because it allows the visualization of fluorescently labelled tumor cells inducing vessel remodeling, disseminating and invading surrounding tissues in a whole transparent embryo. The embryo model has also been used to address the contribution of the tumor stroma in sustaining tumor angiogenesis and spreading. Simultaneously, new anti-angiogenic drugs and compounds affecting malignant cell survival and migration can be tested by simply adding the compound into the water of living embryos. Therefore the zebrafish model offers the opportunity to gain more knowledge on cancer angiogenesis and metastasis in vivo with the final aim of providing new translational insights into therapeutic approaches to help patients.

Wei, G., et al. (2017). "A linear assignment method for multiple criteria decision analysis with hesitant fuzzy sets based on fuzzy measure." International Journal of Fuzzy Systems **19**(3): 607-614.

Xue, Y., et al. (2017). "The Vascular Niche Regulates Hematopoietic Stem and Progenitor Cell Lodgment and Expansion via klf6a-ccl25b." Dev Cell **42**(4): 349-362 e344.

In mammals, hematopoietic stem and progenitor cells (HSPCs) rapidly expand in the fetal liver (FL), but the underlying mechanism remains unclear. Here, we characterize zebrafish caudal hematopoietic tissue (CHT) and identify an important cellular and molecular mechanism of HSPC expansion. Time-lapse imaging showed that HSPCs localize adjacent to vascular endothelial cells (ECs), and their migration and expansion display caudal vein-specific orientation in the CHT. RNA sequencing and functional analysis identified that an EC-expressed transcription factor, Kruppel-like factor 6a (Klf6a), is essential for the CHT niche. We further demonstrated that Klf6a directly regulates the expression of the chemokine (C-C motif) ligand 25b to modulate HSPC lodgment and proliferation. Ex vivo culture results support the conserved role of Ccl21/Ccr7 signaling in promoting HSPC expansion in mammals. Together, we identify the Klf6a-

Ccl25b/Ccr7 axis in controlling the complex HSPC-CHT niche interaction, which may be applicable to in vitro expansion or engraftment of HSPCs after transplantation.

Zaritsky, A., et al. (2021). "Interpretable deep learning uncovers cellular properties in label-free live cell images that are predictive of highly metastatic melanoma." Cell Syst.

ZFIN (2021). "fli1." 1. Retrieved 8/1/2021, 2021, from https://zfin.org/ZDB-GENE-980526-426#summary.