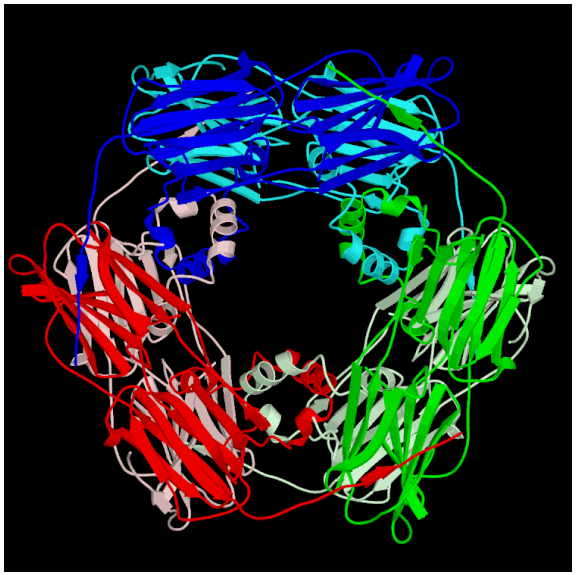
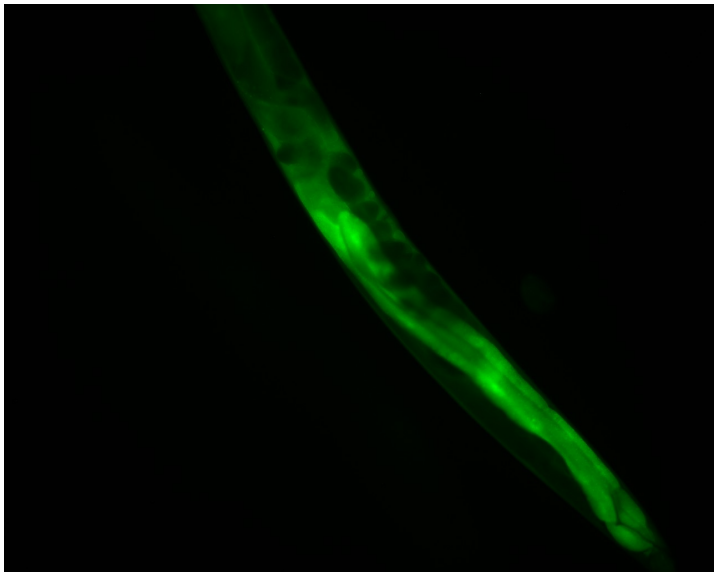
Intelligent Biomedical Data Analysis





Dr. Alexander K. Seewald

SEEWALD SOLUTIONS

Intelligent Biomedical Data Analysis (1)

Involves **Biological & Medical Data**

- Images (2/3D); Text (Research papers, lab notes, DNA); Multi-dimensional „classical“ data (EEG, clinical trials, patient records…)

Involves **Data Analysis**

- Applying statistical tests; parameter fitting to full or partial known models; exploratory data analysis (hypothesis-generating) etc..

Involves **Intelligence**

- Extensive computer use; Analytics which would otherwise be impossible or too expensive.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Intelligent Biomedical Data Analysis (2)

Why is this a challenging field? Because the results of biomedical research are usually not available in computer-readable form!

- Biochemical pathway reaction constants and relations are not present in a formal language.
- Main communication between researchers is in the form of research papers (text), which can only be rudimentarily understood by computers.
- Most online databases are designed to be human-readable rather than computer-readable.
- Image features easily discernable to a human observer are quite hard to teach computers.

...

alex@seewald.at
http://www.seewald.at

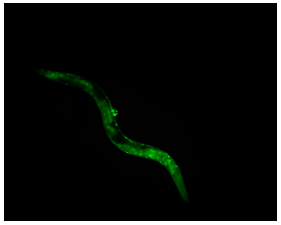
Watching C. Elegans Think (1)

Basic research project in Systems Neuroscience (in cooperation w/ Univ. of Colorado, Boulder)

Four Objectives

- Engineering — *Real-time tracking nerve cells*
- Methodological — *Validate nervous cell models*
- Holistic — *Understand complete N.S.*
- Insight — *Better learning algorithms*

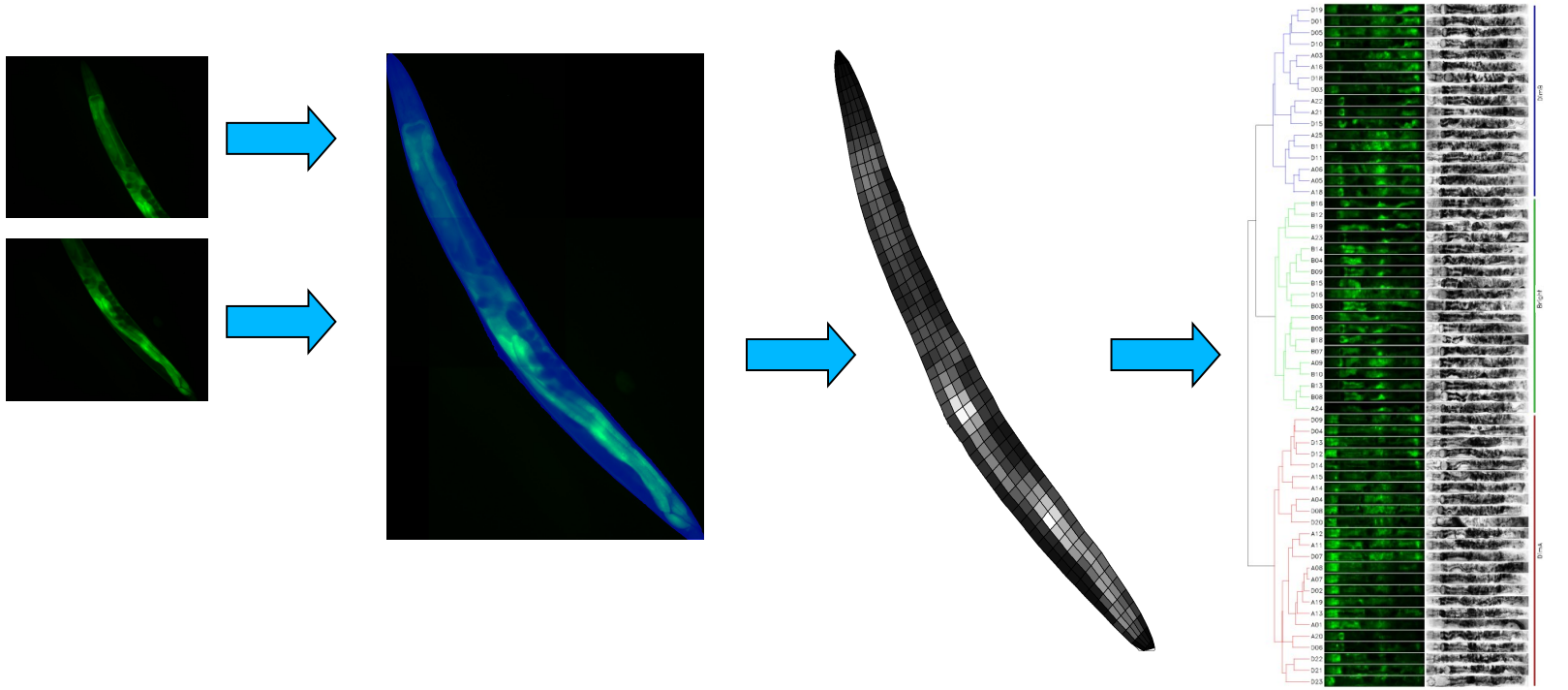
Model organism: C. elegans
~ 1000 cells, ~ 300 nerve cells
Might be feasible to simulate



alex@seewald.at
http://www.seewald.at

Watching C. Elegans Think (2)

Results of an automated analysis of C.elegans



Seewald AK, Cypser J, Mendenhall A, Johnson T (2010) *Quantifying Phenotypic Variation in Isogenic Caenorhabditis elegans Expressing Phsp-16.2::gfp by Clustering 2D Expression Patterns*, PLoS ONE 5(7): e11426. doi:10.1371/journal.pone.0011426.

Prototype GPL source code: http://elegans.seewald.at/

alex@seewald.at
http://www.seewald.at

Watching C. Elegans Think (3)

Quantifying Phenotypic Variation...

Analyzing changes in appearance / phenotype...

in Isogenic Caenorhabditis elegans...

in small nematodes (worms) which all have the same genetic code (i.e. clones)...

Expressing Phsp-16.2::gfp...

which express a GFP reporter that binds to heat shock protein 16 (i.e. transgenic worms)...

by Clustering 2D Expression Patterns

by extracting 2D expression patterns that are independent of worm pose AND clustering these patterns using hierarchical clustering methods

alex@seewald.at
http://www.seewald.at

Watching C. Elegans Think (4)

Heat Shock Protein 16 – increases expression e.g. when organism is exposed to high temperatures

HSP are named by molecular weight (=16kD). Expressed in intestine and pharynx. Induced in response to heat shock or other environmental stresses

Interacts with intracellular human beta amyloid peptide (Alzheimer plagues)

High expr. correlates with worm longevity acc. to earlier studies.

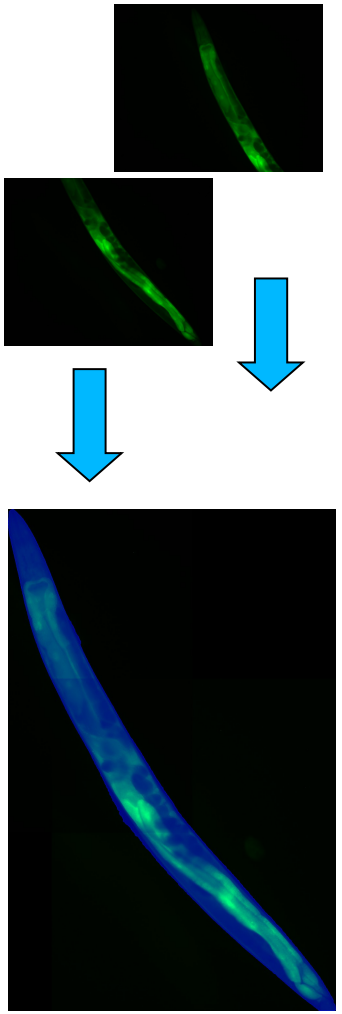
alex@seewald.at
http://www.seewald.at

Image Processing Methods (1)

Merging anterior/posterior image & extracting worm

(pixelClassification)

- Machine learning from manually tagged sample images (i.e. „Ground Truth“)
- Threshold optimization by testing minimum circularity and area of largest blob
- Closure (erode, dilate), fill internal holes with circularity below threshold
- Heuristic search for breaks in contour, which are repaired with straight lines and filled on the inside



(meshAB)

- Image correlation coefficient for combining head and tail images (simplified stitching)

alex@seewald.at
http://www.seewald.at

Image Processing Methods (2)

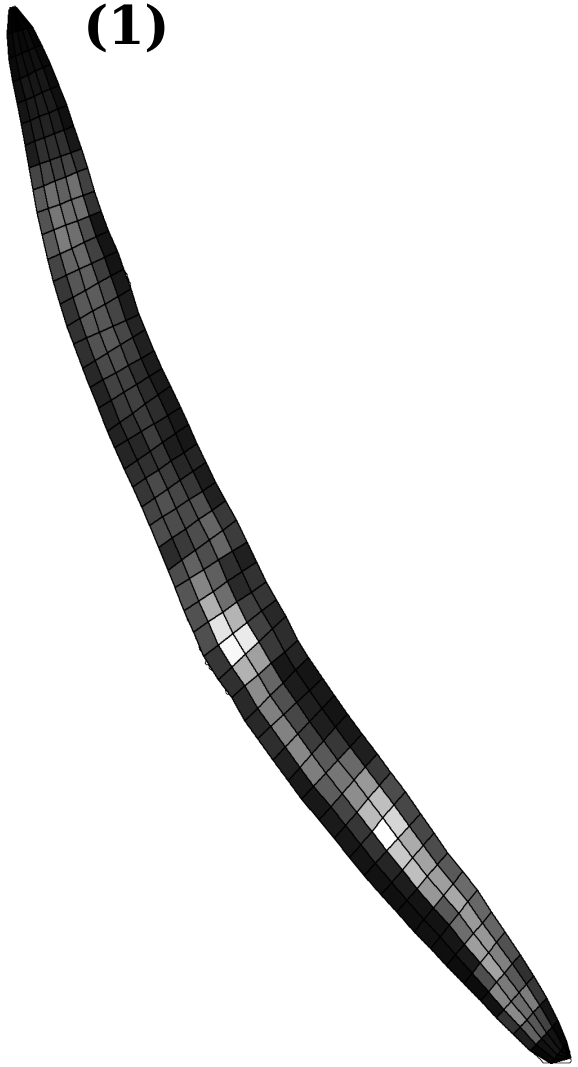
Computing 2D expression patterns (1)

(meshAB)

- Computing head/tail/vulva pos. from worm pose via heuristics (~ 50%)
- Manual markup in remaining cases

(sampleCE)

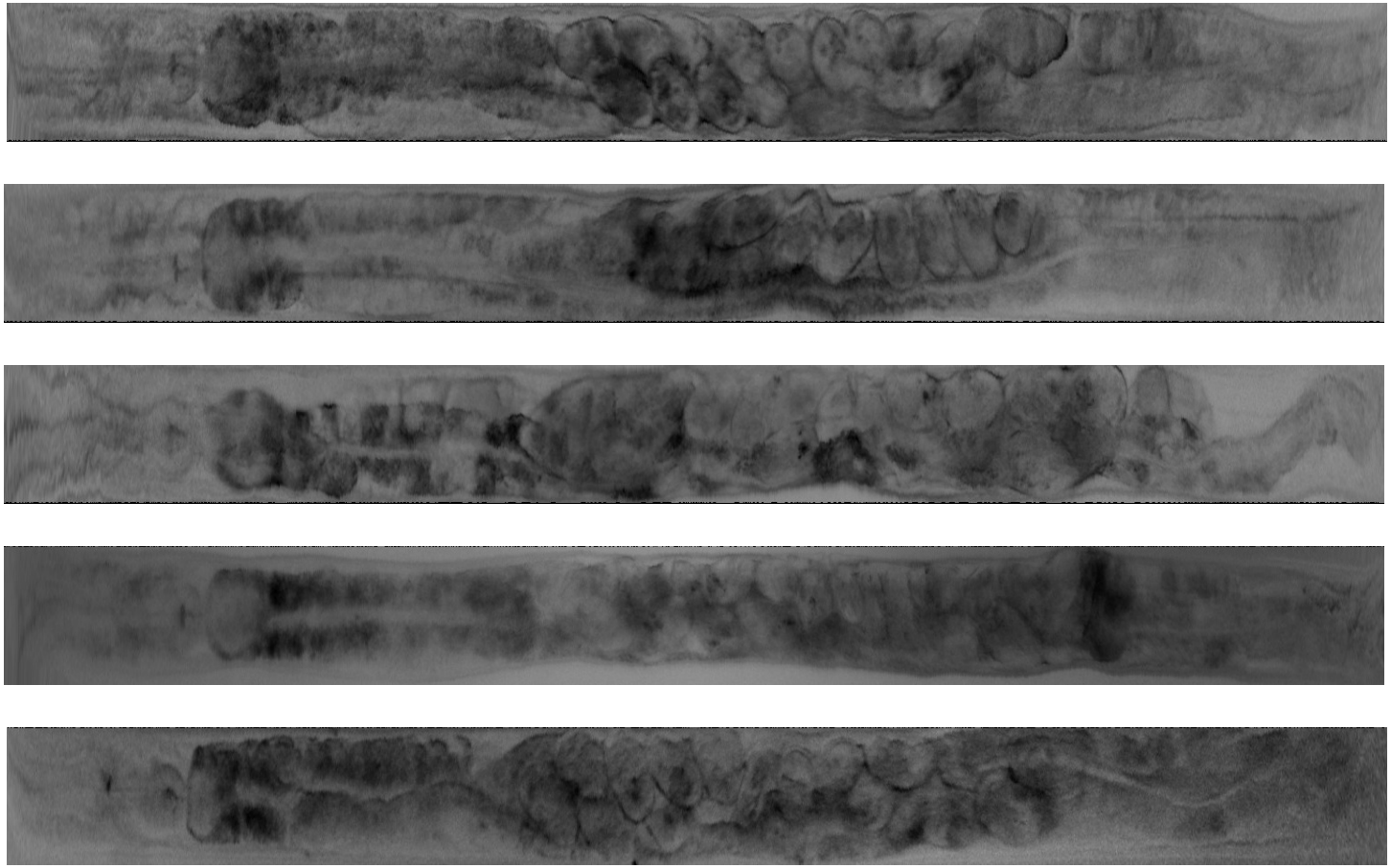
- Worm backbone via „thinning“
- Search left/right from backbone in perpendicular direction to local curvature for worm border
- Split left/right and top/bottom into the desired number of tiles
- Comp.avg.GFP intensity per tile



alex@seewald.at
http://www.seewald.at

Image Processing Methods (3)

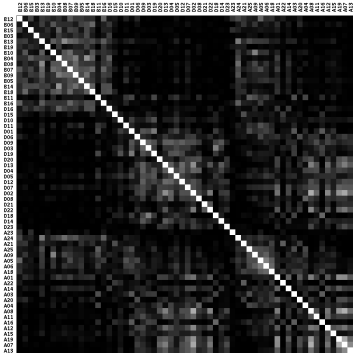
Computing 2D expression patterns (2)

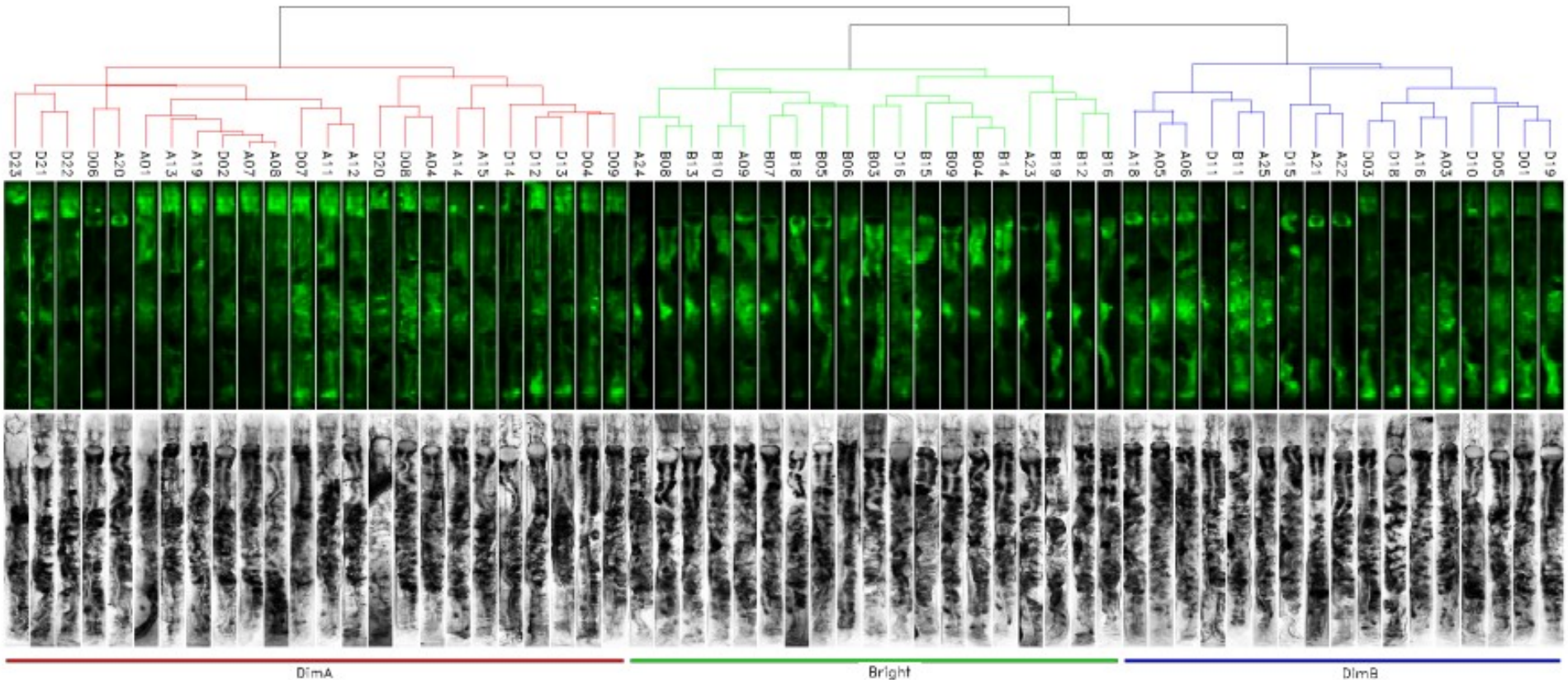


alex@seewald.at
http://www.seewald.at

Results (1)

- Bright worms live longer than dim ones
- Even when discounting brightness, bright worms show distinct expression patterns (currently under investigation)

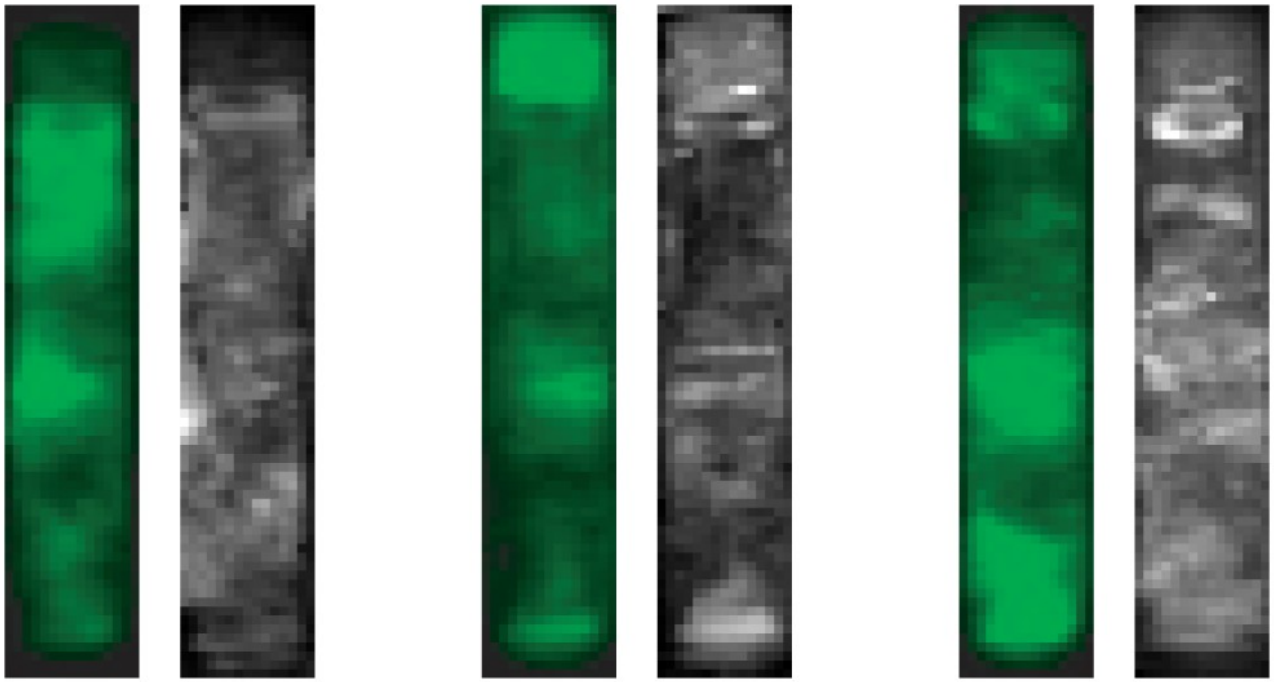




Results (2)

- Bright worms: one cluster
- Dim worms: two clusters w/ distinct expr. patterns

Bright **DimA** **DimB**

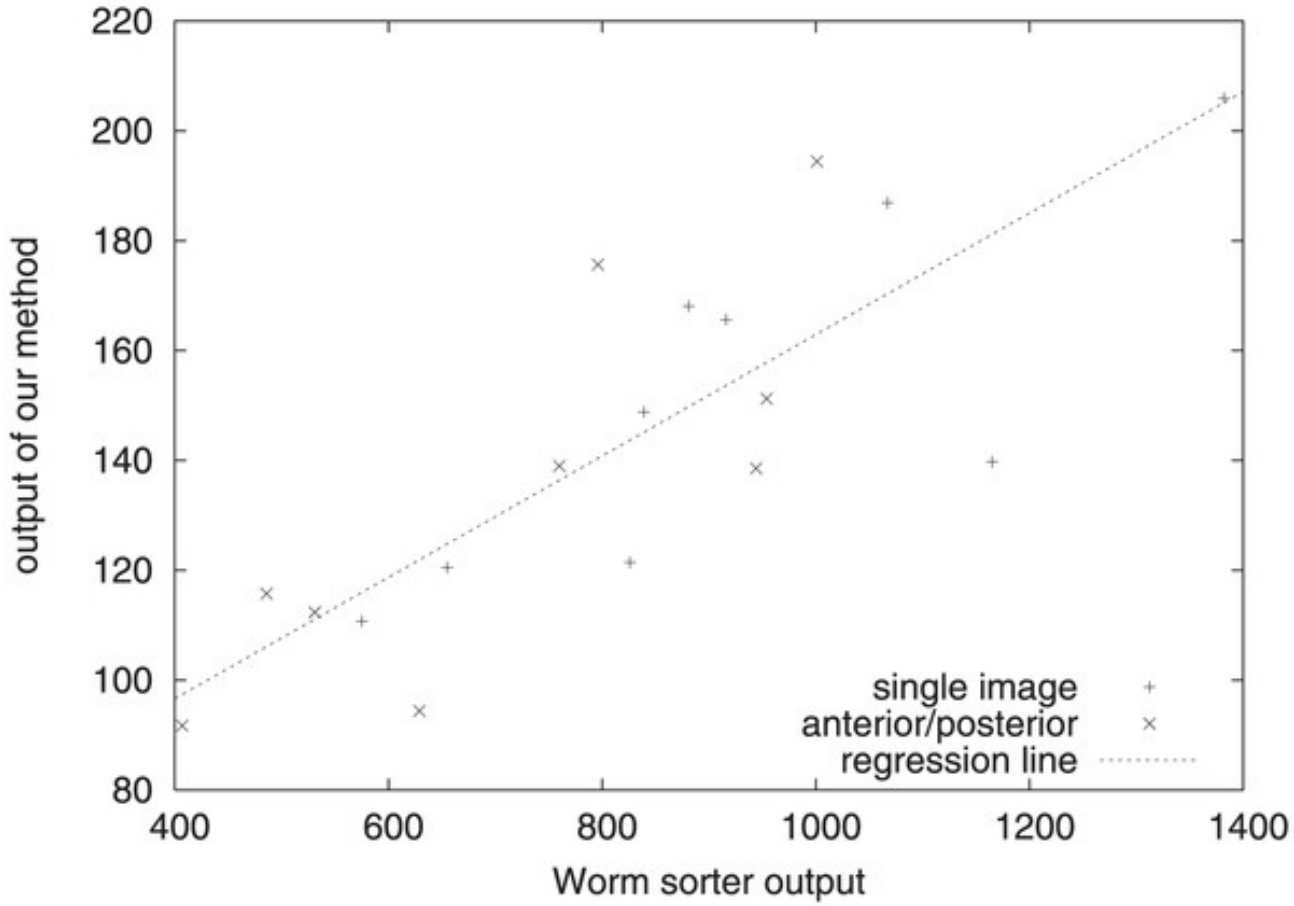


Top: head, bottom: tail, vulva to the right.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Results (3)

Validation versus worm sorter: very good agreement!

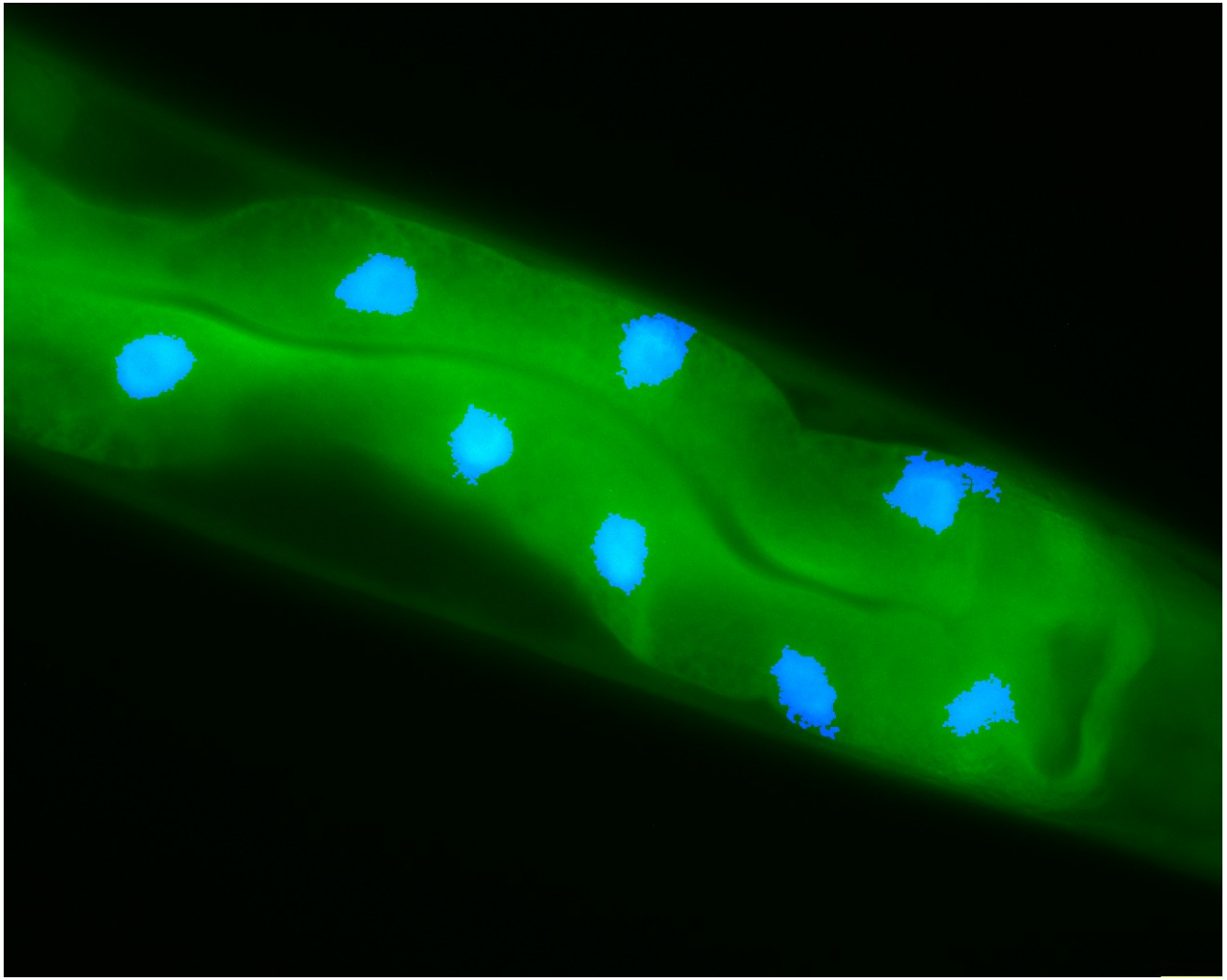


SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

Current Work

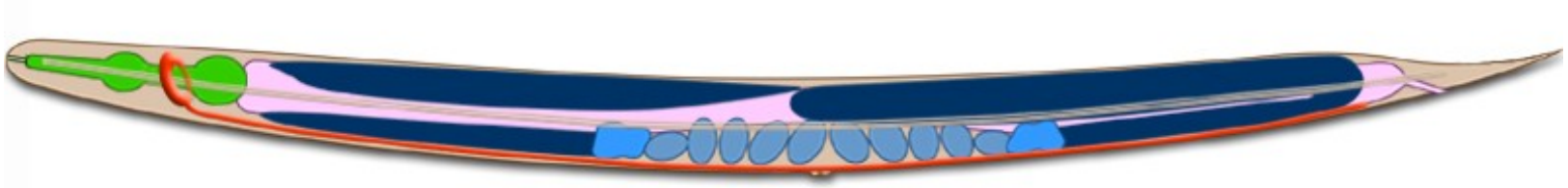
Nuclei detection for high-resolution images



SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

Future Work (1)



- Bottleneck is image acquisition – each worm has to be taken from culture medium, anesthesized, cleaned and imaged (ca. 30min per worm)
- Resolution is too coarse for observing single cells

→ *Culturing worms on chamber-slides, using slide-based microscopy & automated imaging*

- Lots of problems with different microscope settings, air bubbles, finetuning,...

→ *„Closed-loop“ system (microscope, moveable slide and image analysis in a coupled system)*

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

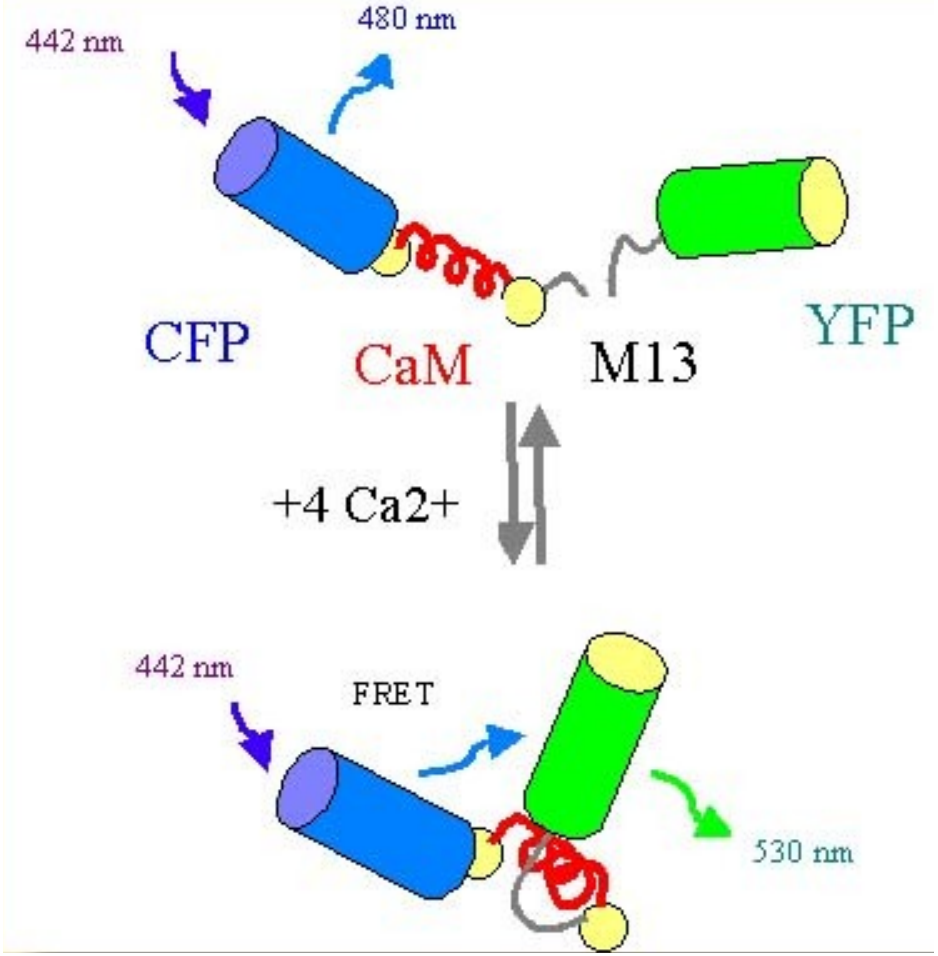
Future Work (2)

Cameleon: measures Ca2+ level = nerve cell activity

CFP emits a 480nm photon on excitation with 442nm.

High Ca2+ concentrations lead to conformation changes and the photon is absorbed by YFP and re-emitted as 530nm.

Proportion between 480nm and 530nm response used as signal for Ca2+ level.



alex@seewald.at
http://www.seewald.at