Machine Learning Supported Interactive Visualization of Hybrid 3D and 2D Data for the Example of Plant Cell Lineage Specification

INVENTEURS DU MONDE NUMÉRIQUE

université PARIS-SACLAY

LISN LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

NAVISCOPE

Thèse soutenue à Paris-Saclay, le 14 février 2023, par

Jiayi HONG

Supervisée par Tobias Isenberg et Alain Trubuil

Composition du Jury

Miriah Meyer, *Linköping University*	Rapportrice
Barbora Kozlíková, *Masaryk University*	Rapportrice
Caroline Appert, *Inria & Univ Paris-Saclay*	Examinatrice
Anastasia Bezerianos, *Inria & Univ Paris-Saclay*	Examinatrice
Johanna Beyer, *Harvard University*	Examinatrice
Tobias Isenberg, *Inria & Univ Paris-Saclay*	Directeur
Alain Trubuil, *INRAE*	Co-directeur

Machine Learning Supported Interactive Visualization of Hybrid 3D and 2D Data for the Example of Plant Cell Lineage Specification

Thèse soutenue à Paris-Saclay, le 14 février 2023, par

Jiayi HONG

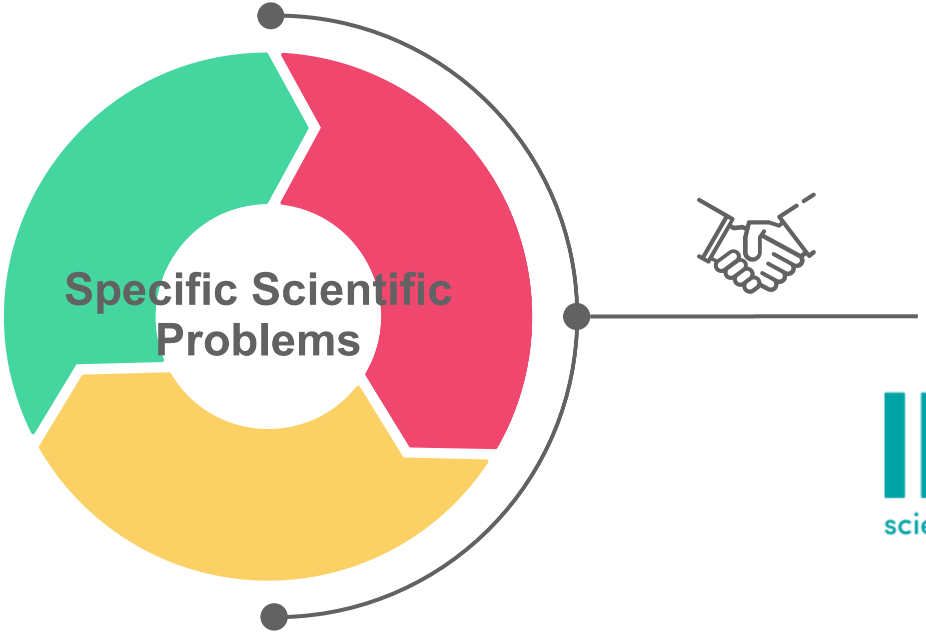
Supervisée par Tobias Isenberg et Alain Trubuil

Composition du Jury

Miriah Meyer, *Linköping University*	Rapportrice
Barbora Kozlíková, *Masaryk University*	Rapportrice
Caroline Appert, *Inria & Univ Paris-Saclay*	Examinatrice
Anastasia Bezerianos, *Inria & Univ Paris-Saclay*	Examinatrice
Johanna Beyer, *Harvard University*	Examinatrice
Tobias Isenberg, *Inria & Univ Paris-Saclay*	Directeur
Alain Trubuil, *INRAE*	Co-directeur

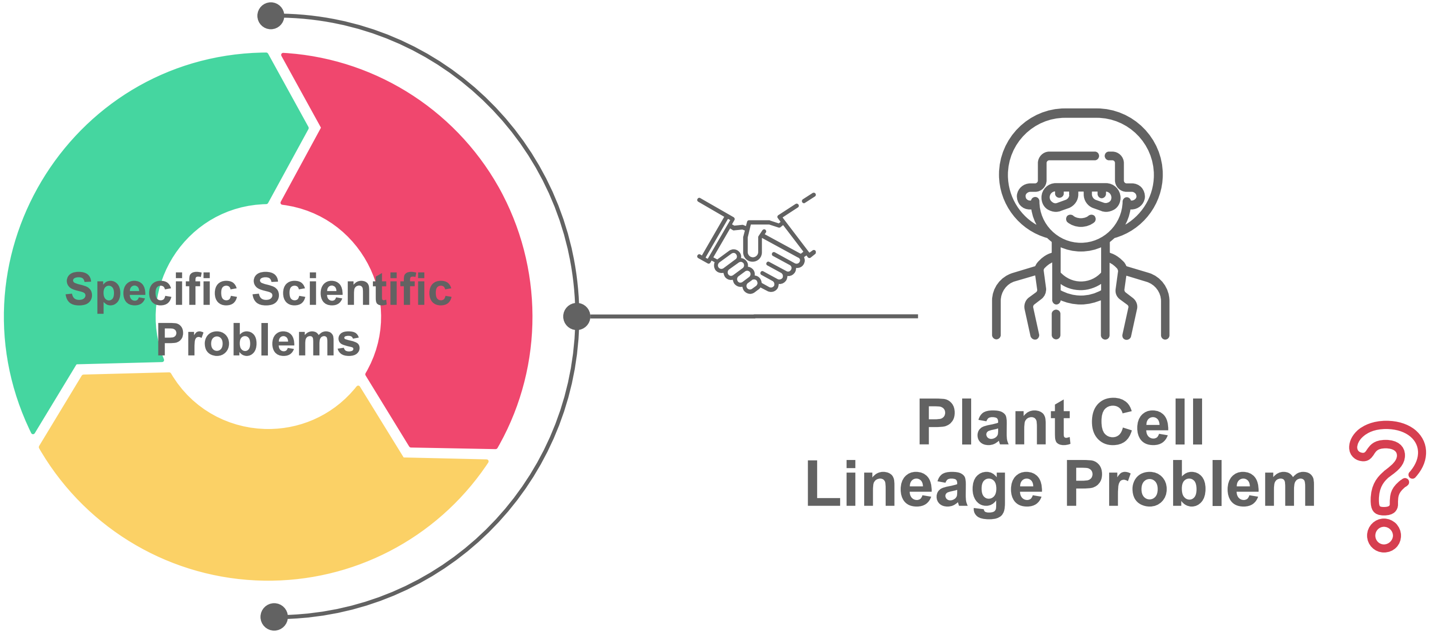
Research Scope

Research Scope



INRAE
science for people, life & earth

Research Scope

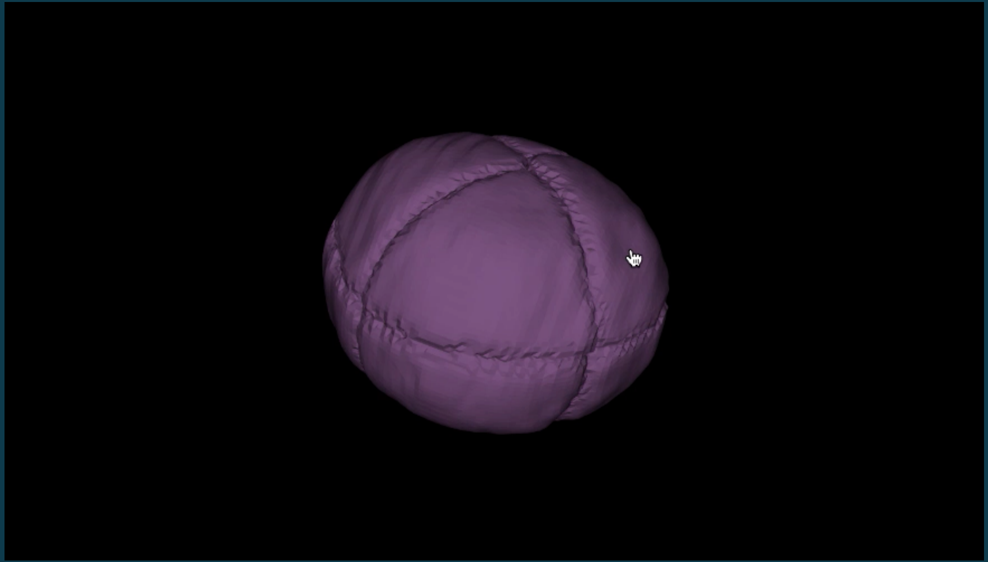


01

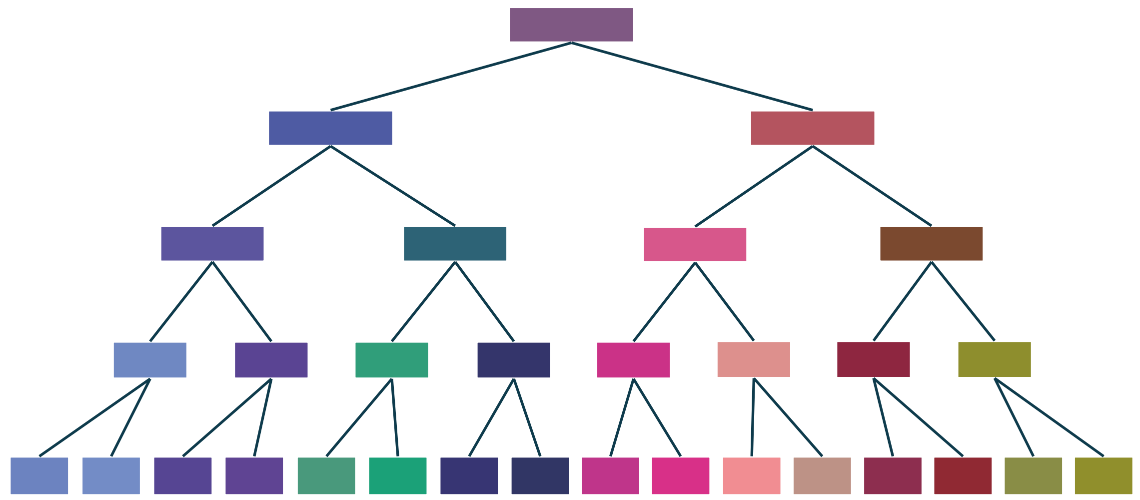
Background

Plant Cell Lineage

The development history of plant embryos.

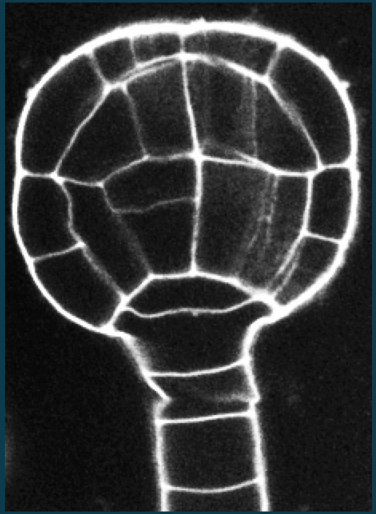


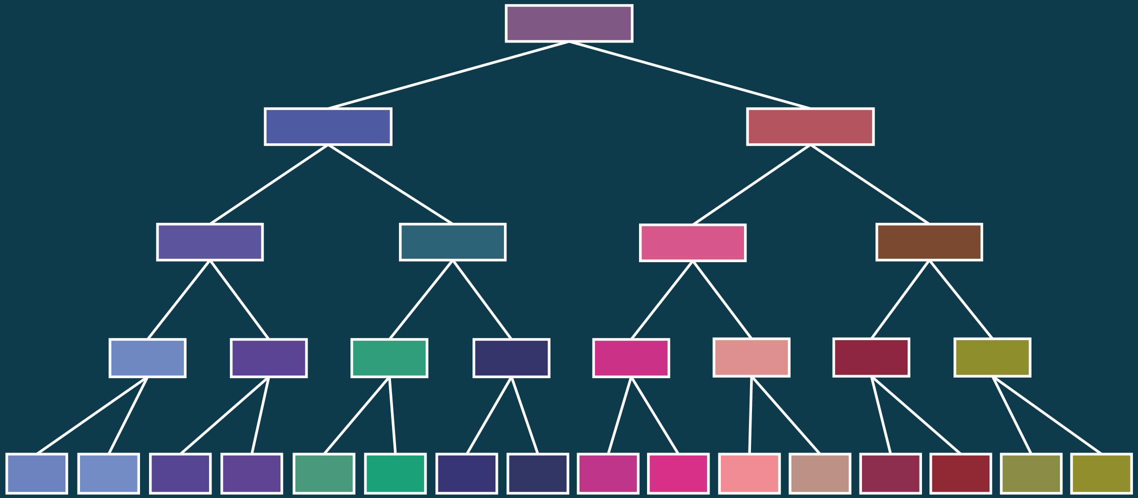
Plant Cell Lineage



Plant Cell Lineage

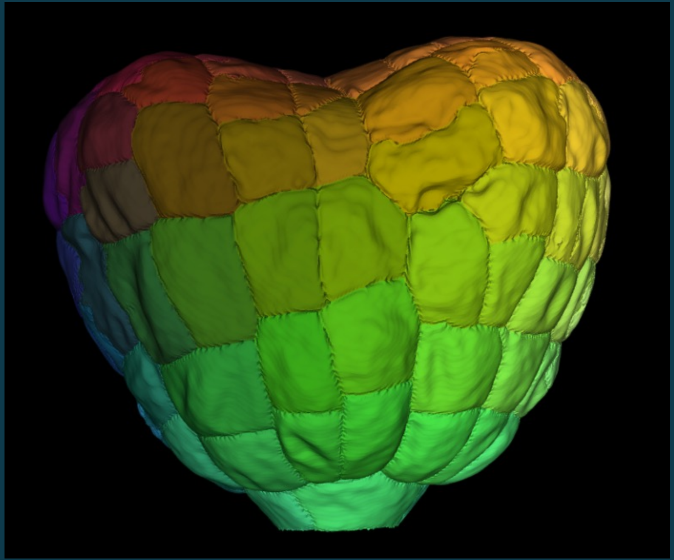
For a specific embryo, what do biologists do to build the hierarchy?

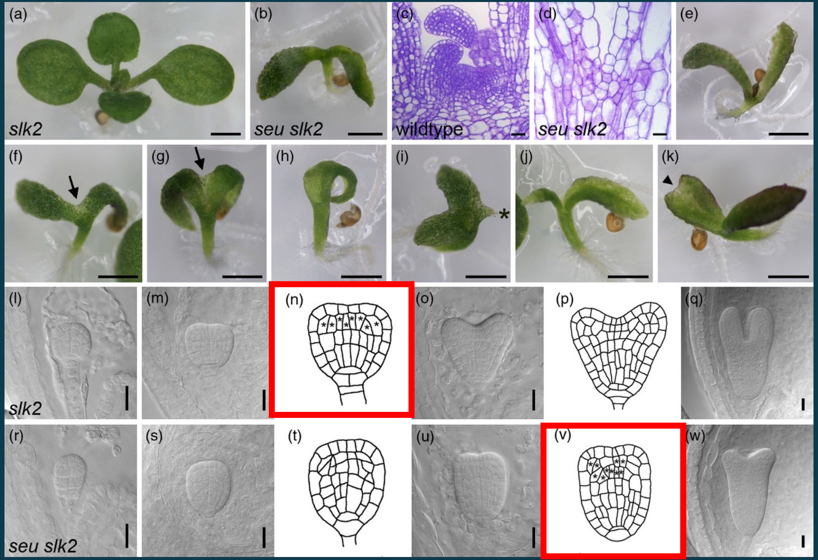




Purpose of Plant Cell Lineage

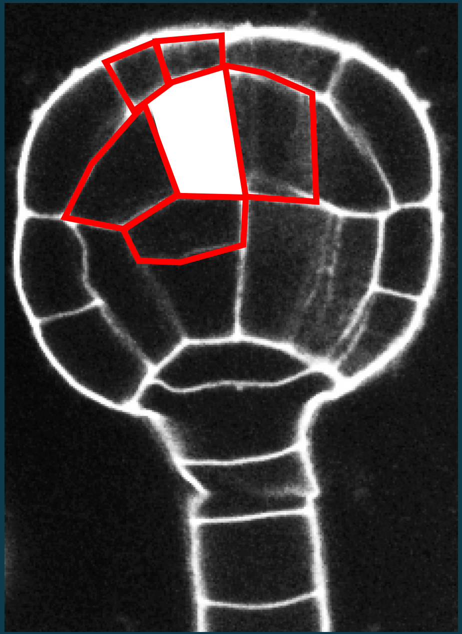
Why should biologists investigate cell lineage?





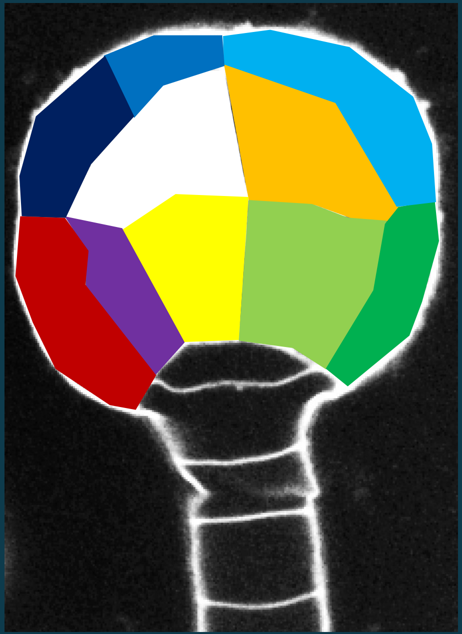
[Lee, et al. 2014]

How to do cell lineage?



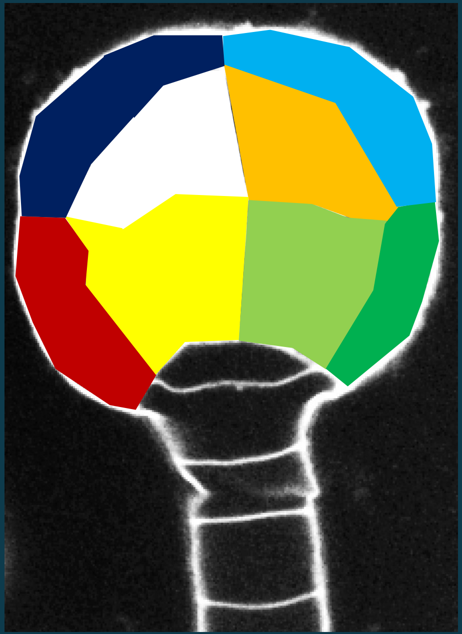
Biologists need to find the right sister cell for every cell in an embryo.

How to do cell lineage?



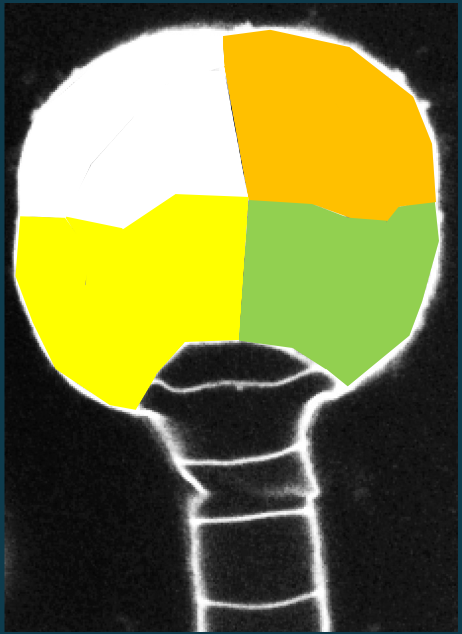
Once decided, they would merge cells and continue assigning the remaining cells.

How to do cell lineage?



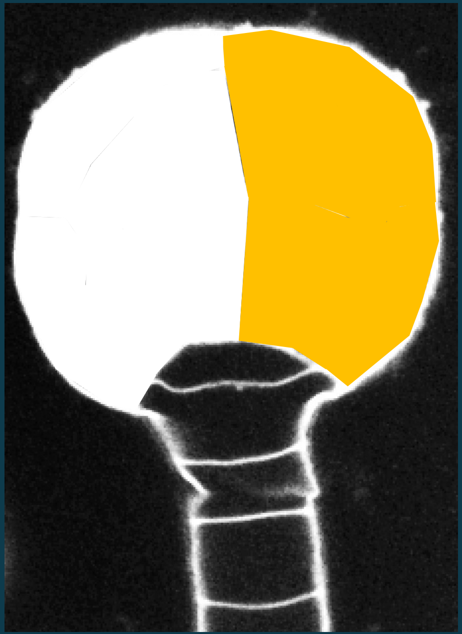
Biologists will continue this process to the new generation until there is only one cell left.

How to do cell lineage?



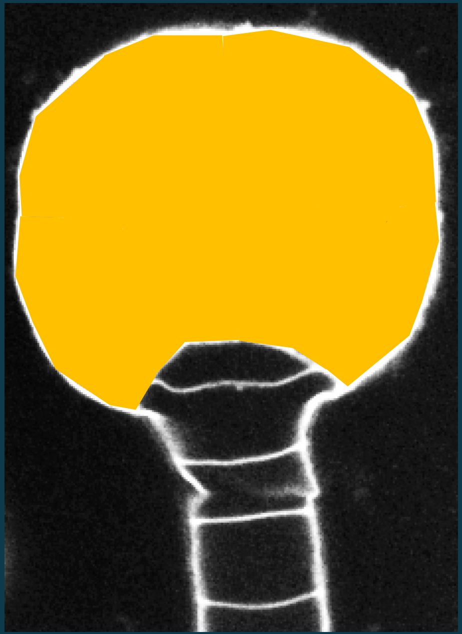
Biologists will continue this process to the new generation until there is only one cell left.

How to do cell lineage?



Biologists will continue this process to the new generation until there is only one cell left.

How to do cell lineage?



Biologists will continue this process to the new generation until there is only one cell left.

Traditional Workflow

Get segmented 2D slices

Check every cell

Manually find the sister for each cell

Write down the hierarchy on paper

Challenges

01

No Absolute Truth

The embryo has already divided.

02

Decision-making

It relies on biologists' experience and knowledge.

03

Time-consuming

An embryo can be large.

04

Tedious

The cell assignment is repetitive.

Traditional Workflow

Get segmented 2D slices

Check every cell

Manually find the sister for each cell

Write down the hierarchy on paper

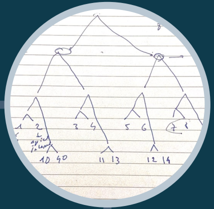
Traditional Workflow

Get segmented 2D slices

Interaction

Manually find the sister for each cell

Visualization



Visualization

Check every cell

Visualization

Interaction

ML

Write down the hierarchy on paper

02

First Project

Traditional Workflow

Get segmented 2D slices

Visualization

Interaction

Check every cell

Manually find the sister for each cell

Visualization

Interaction

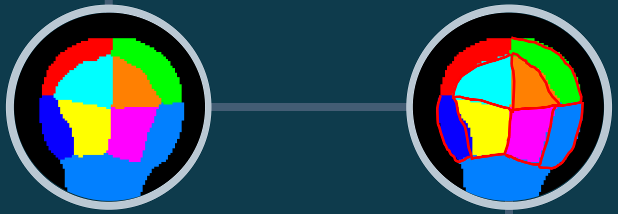
ML

Visualization

Write down the hierarchy on paper

Get segmented 2D slices

Interaction



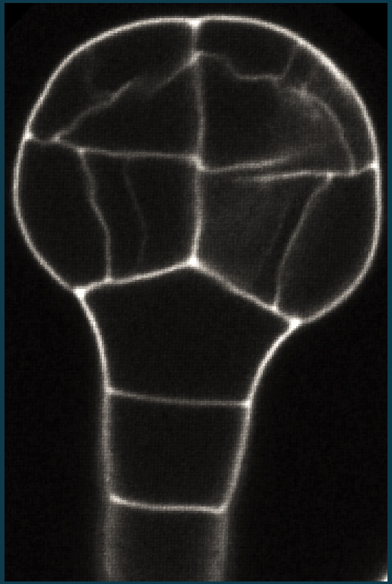
Visualization

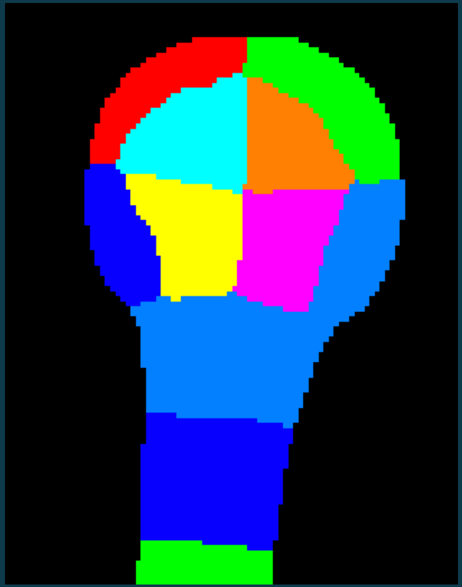
Check every cell

The Problem of Plant Cell Embryos

Difficult to have an overview of each cell.

Visualization

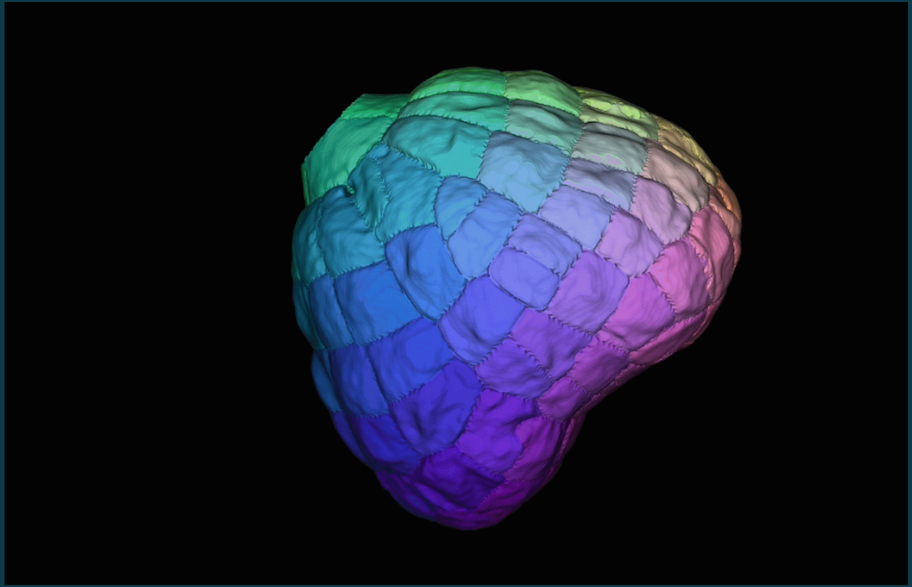




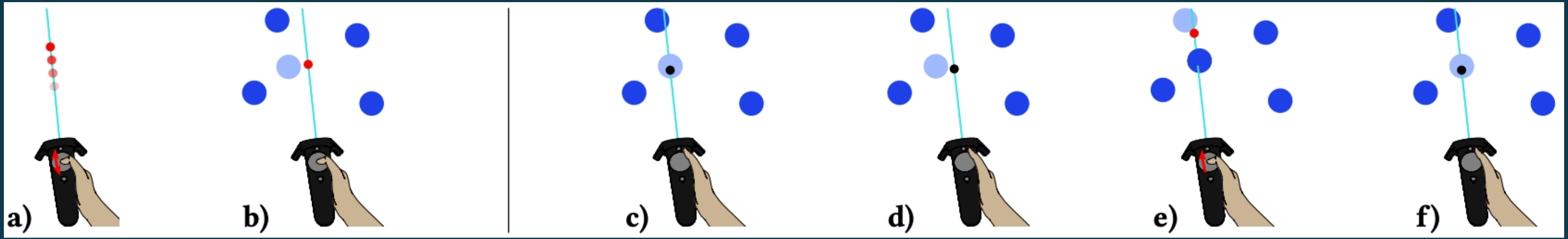
The Problem of Plant Cell Embryos

Cells are densely packed together.

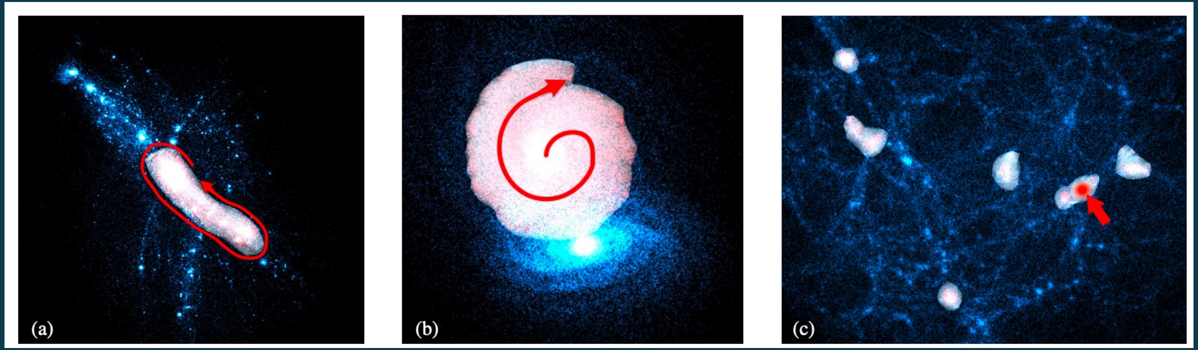
Interaction



Related Work

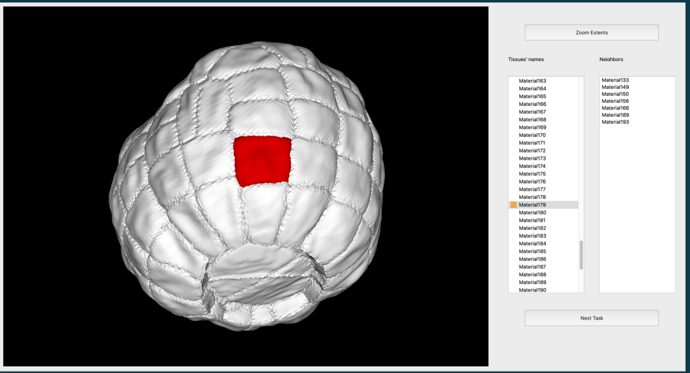


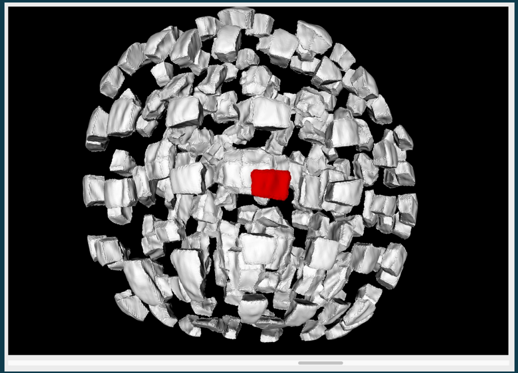
[Baloup et al. 2019]

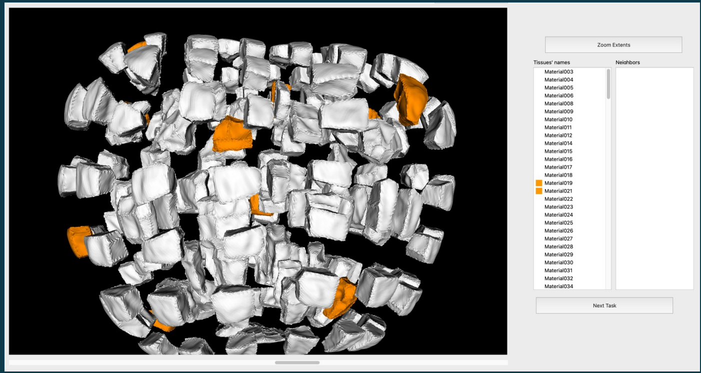


[Yu et al. 2016]

Exploration







1D

Select from the list

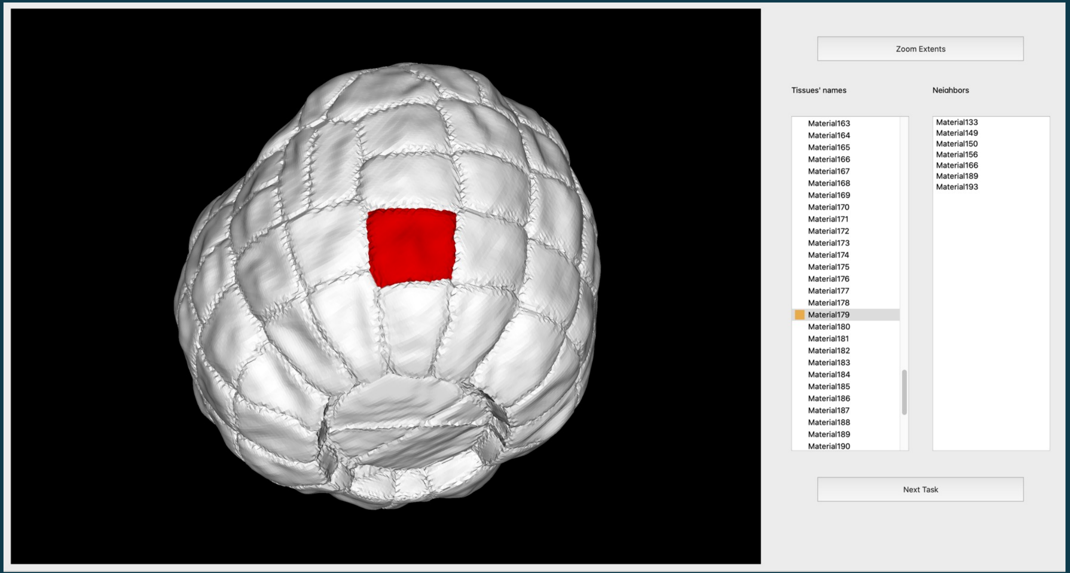
3D

Select from the explosion view

1D + 3D

Select from both views

List Selection

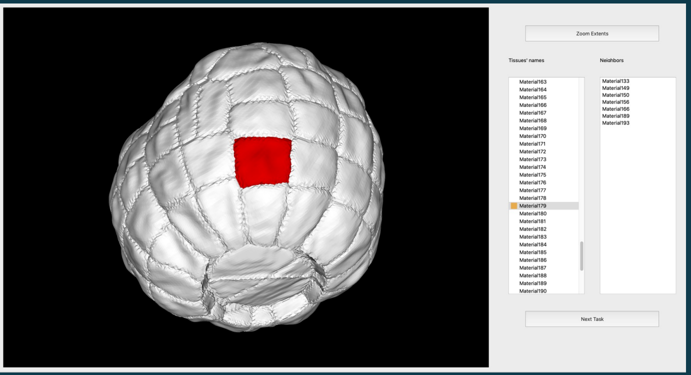


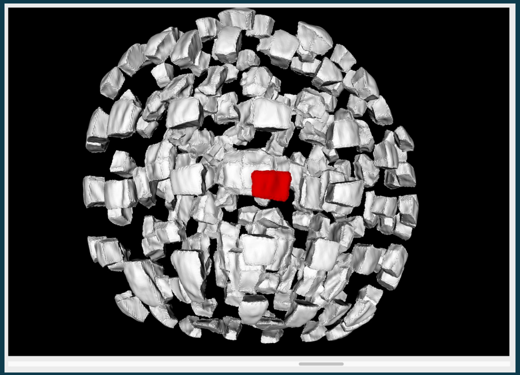
1D

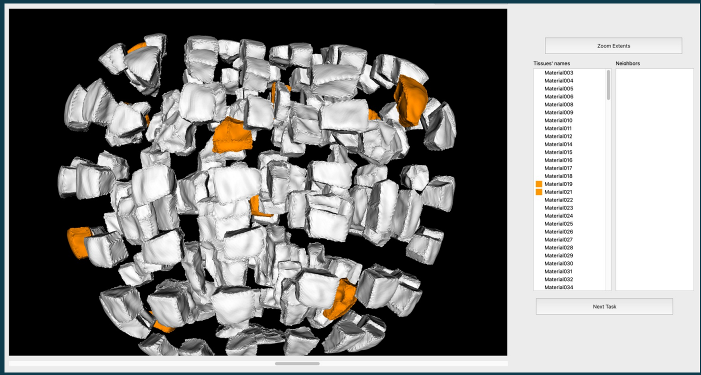
Select from the list

Use the list to traverse all the cells and their neighbors. Also, use the list to get access to cells.

Exploration

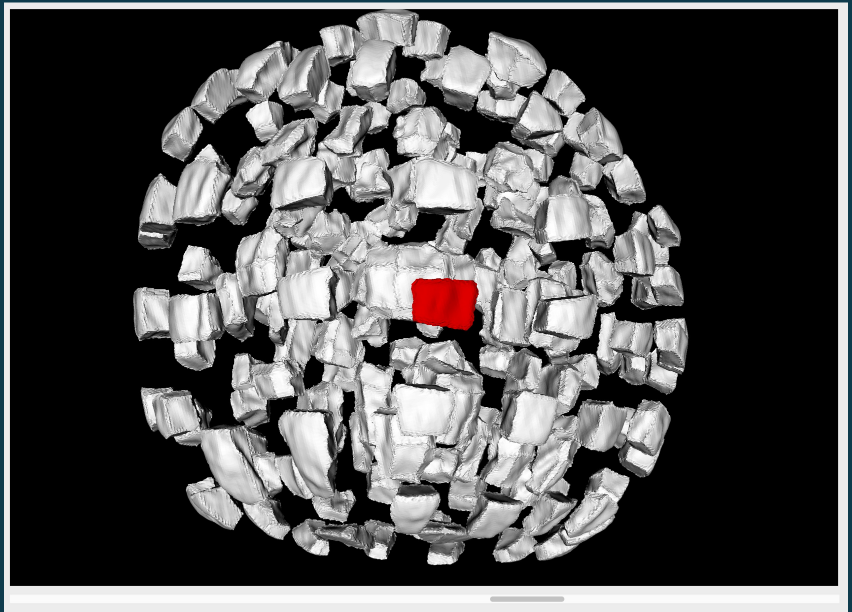






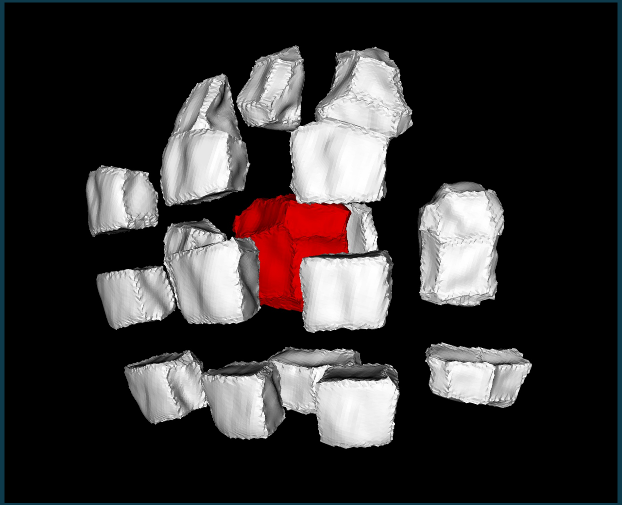
1D	3D	1D + 3D
Select from the list	Select from the explosion view	Select from both views

Explosion Selection



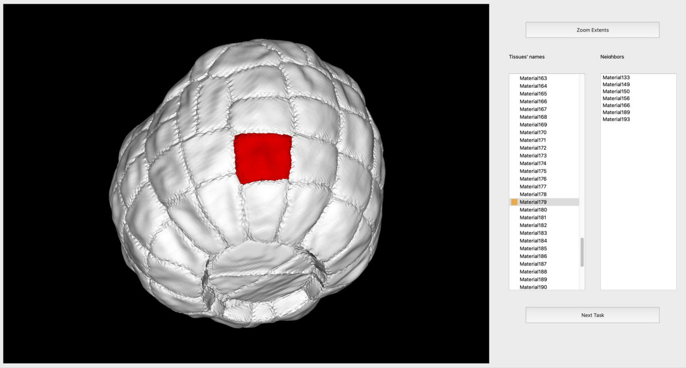
3D

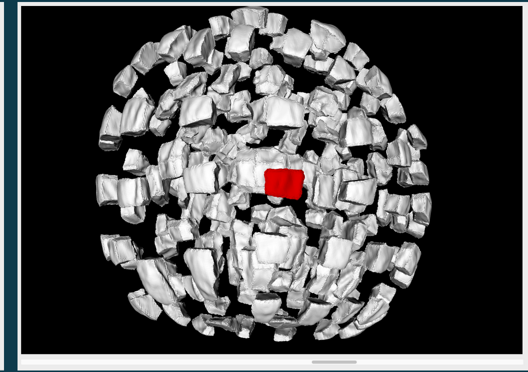
Select from the explosion view

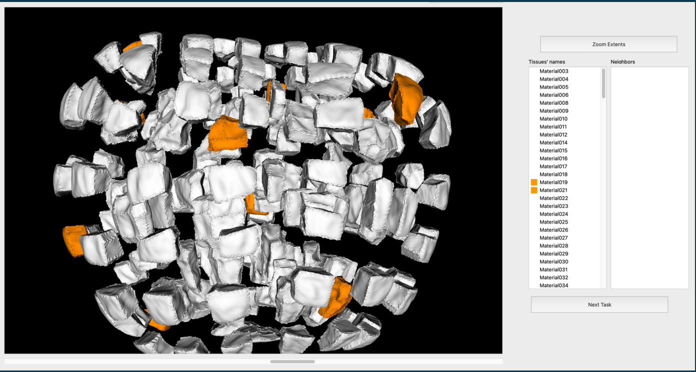


People traverse neighboring cells in 3D and get access to cells in the explosion view.

Exploration







1D

Select from the list

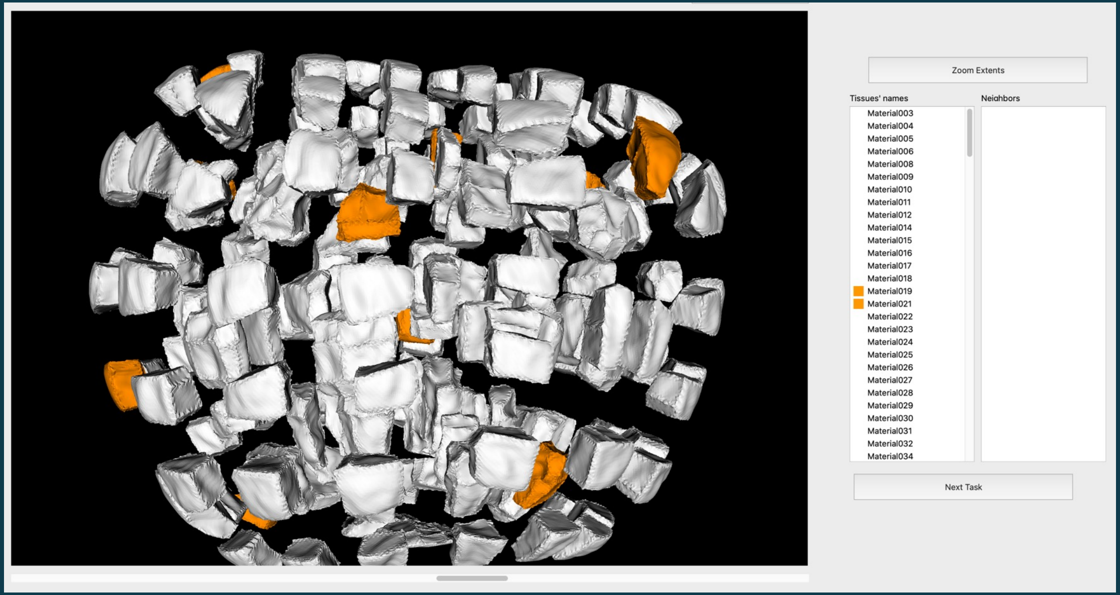
3D

Select from the explosion view

1D + 3D

Select from both views

Combination Selection



1D + 3D

Select from both views

People can traverse and access the cells of an embryo in both ways.

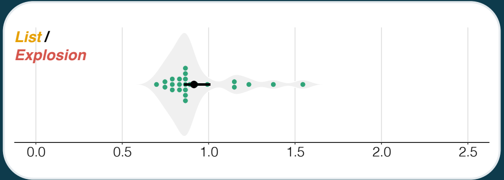
User Study

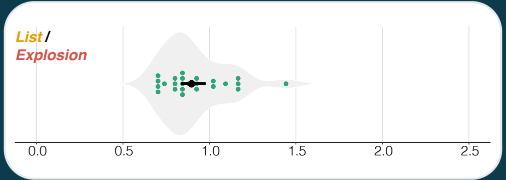
Overall

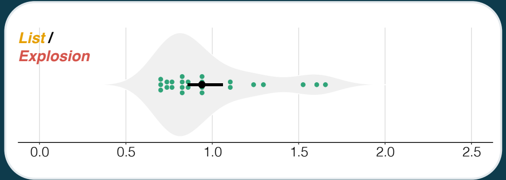
Inside Cells

Outside Cells

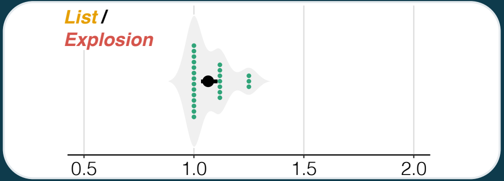
Completion Time

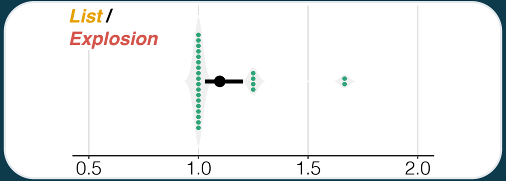


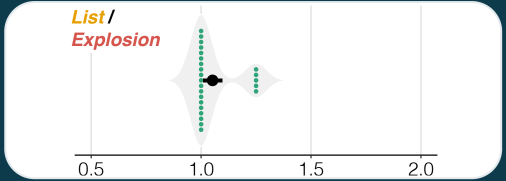




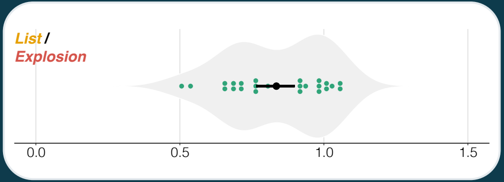
Accuracy

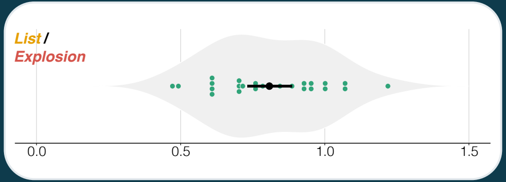


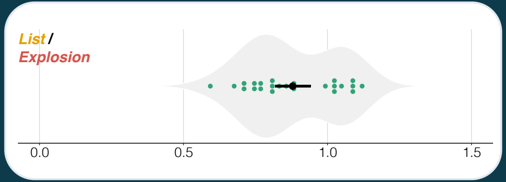




Clicking Ratio







Study Results

- The explosion technique makes it possible to access objects in such a tightly packed 3D environment.
- Combination (List + Explosion) seems to combine these advantages of the single technique.

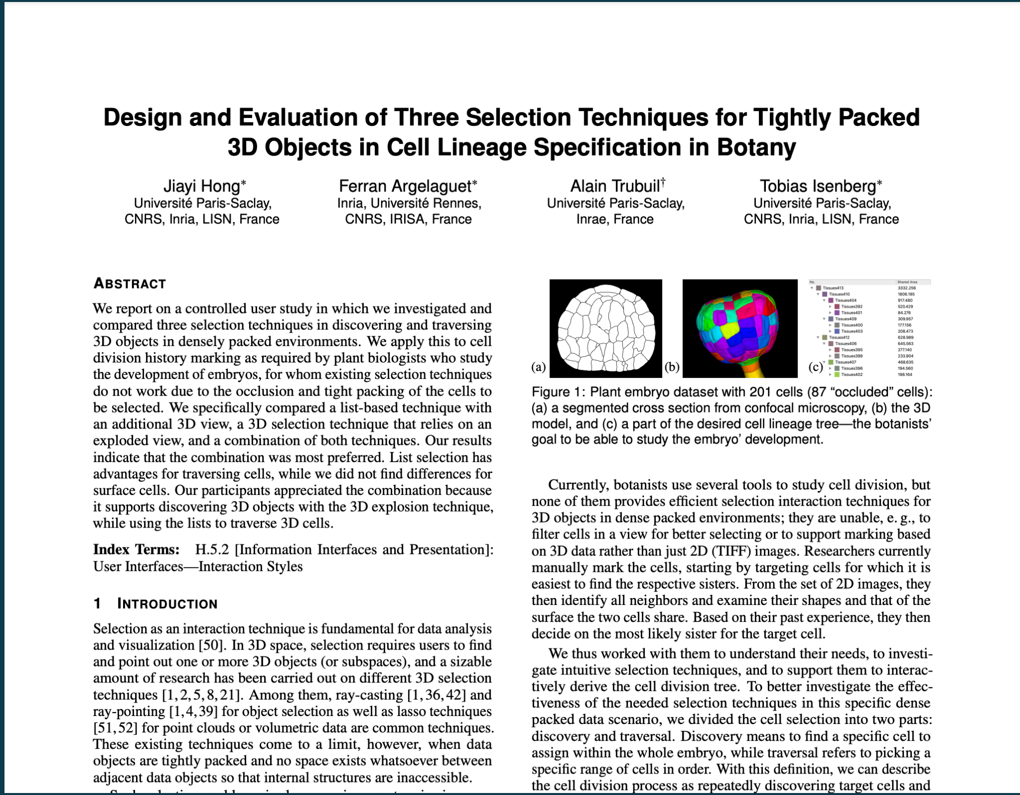
Publication

Jiayi Hong, Ferran Argelaguet, Alain Trubuil, and Tobias Isenberg. "Design and Evaluation of Three Selection Techniques for Tightly Packed 3D Objects in Cell Lineage Specification in Botany".

In Proceedings of Graphics Interface (GI, May 27–28, virtually in Vancouver, BC, Canada). Mississauga, ON, Canada: Canadian Human-Computer Communications Society, 2021, pp. 213–223.

ISBN: 978-0-9947868-6-9

DOI: 10.20380/GI2021.33



Design and Evaluation of Three Selection Techniques for Tightly Packed 3D Objects in Cell Lineage Specification in Botany

Jiayi Hong* Université Paris-Saclay, CNRS, Inria, LISN, France
Ferran Argelaguet* Inria, Université Rennes, CNRS, IRISA, France
Alain Trubuil† Université Paris-Saclay, Inrae, France
Tobias Isenberg* Université Paris-Saclay, CNRS, Inria, LISN, France

Abstract

We report on a controlled user study in which we investigated and compared three selection techniques in discovering and traversing 3D objects in densely packed environments. We apply this to cell division history marking as required by plant biologists who study the development of embryos, for whom existing selection techniques do not work due to the occlusion and tight packing of the cells to be selected. We specifically compared a list-based technique with an additional 3D view, a 3D selection technique that relies on an exploded view, and a combination of both techniques. Our results indicate that the combination was most preferred. List selection has advantages for traversing cells, while we did not find differences for surface cells. Our participants appreciated the combination because it supports discovering 3D objects with the 3D explosion technique, while using the lists to traverse 3D cells.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction Styles

1 Introduction

Selection as an interaction technique is fundamental for data analysis and visualization [50]. In 3D space, selection requires users to find and point out one or more 3D objects (or subspaces), and a sizable amount of research has been carried out on different 3D selection techniques [1, 2, 5, 8, 21]. Among them, ray-casting [1, 36, 42] and ray-pointing [1, 4, 39] for object selection as well as lasso techniques [51, 52] for point clouds or volumetric data are common techniques. These existing techniques come to a limit, however, when data objects are tightly packed and no space exists whatsoever between adjacent data objects so that internal structures are inaccessible.

(a) (b) (c)

Figure 1: Plant embryo dataset with 201 cells (87 "occluded" cells): (a) a segmented cross section from confocal microscopy, (b) the 3D model, and (c) a part of the desired cell lineage tree—the botanists' goal to be able to study the embryo' development.

Currently, botanists use several tools to study cell division, but none of them provides efficient selection interaction techniques for 3D objects in dense packed environments; they are unable, e. g., to filter cells in a view for better selecting or to support marking based on 3D data rather than just 2D (TIFF) images. Researchers currently manually mark the cells, starting by targeting cells for which it is easiest to find the respective sisters. From the set of 2D images, they then identify all neighbors and examine their shapes and that of the surface the two cells share. Based on their past experience, they then decide on the most likely sister for the target cell.

We thus worked with them to understand their needs, to investigate intuitive selection techniques, and to support them to interactively derive the cell division tree. To better investigate the effectiveness of the needed selection techniques in this specific dense packed data scenario, we divided the cell selection into two parts: discovery and traversal. Discovery means to find a specific cell to assign within the whole embryo, while traversal refers to picking a specific range of cells in order. With this definition, we can describe the cell division process as repeatedly discovering target cells and

Traditional Workflow

Get segmented 2D slices

Interaction

Manually find the sister for each cell

Visualization

Visualization

Check every cell

Visualization

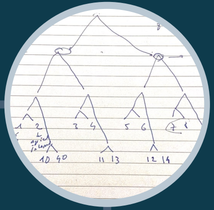
Interaction

ML

Write down the hierarchy on paper

Manually find the sister for each cell

Visualization



Visualization

Interaction

ML

Write down the hierarchy on paper

03

Second Project

LineageD

Datasets ≡

Color Modes

By District By Normalized Shared Area By Model Confidence

Randomized

Explosion Extent

Peeling

Interactions

Confirm | Wrong Children | New Sister | Supporting Cell

Predict One Level

Selections:

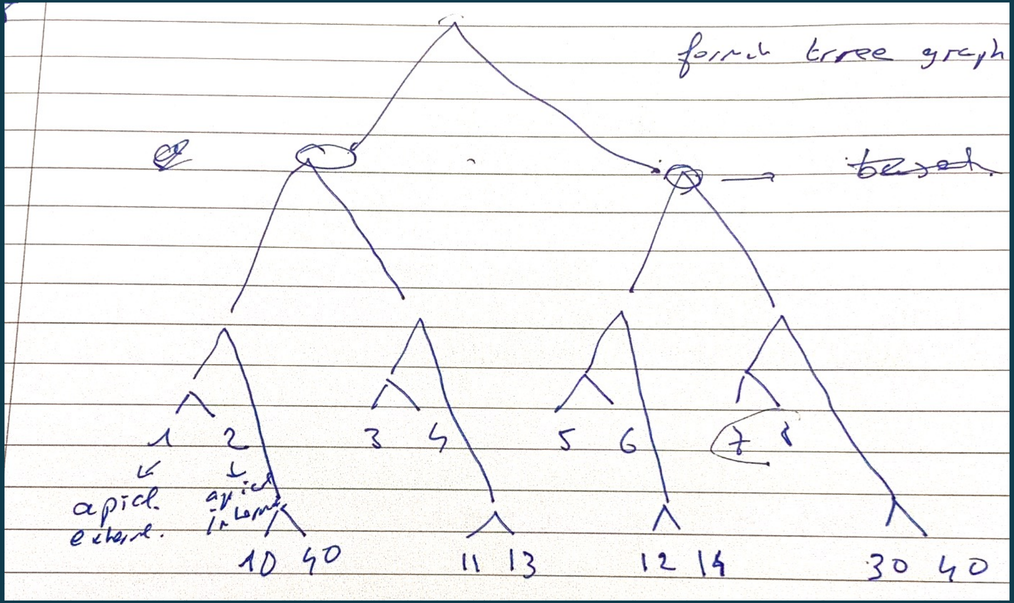
Target & Sister

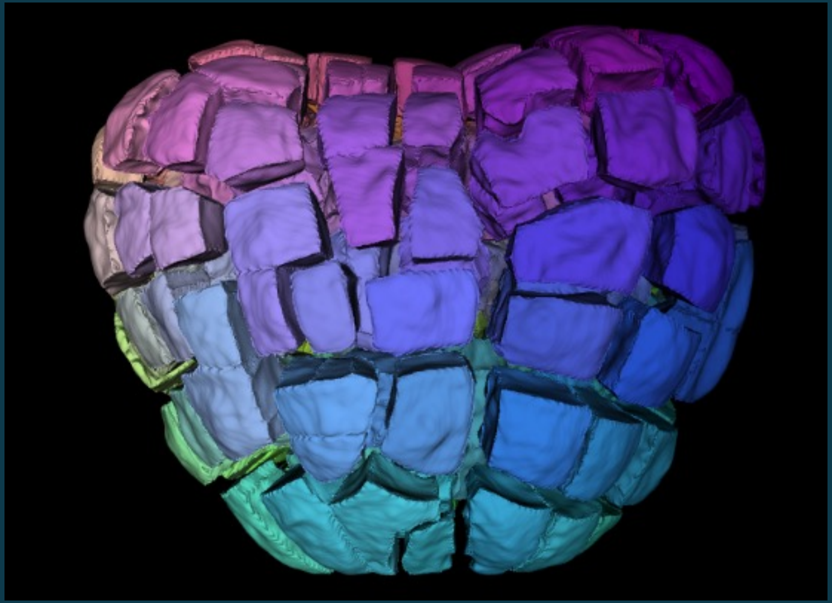
Interactions

Confirm | Wrong Children | New Sister | Supporting Cell

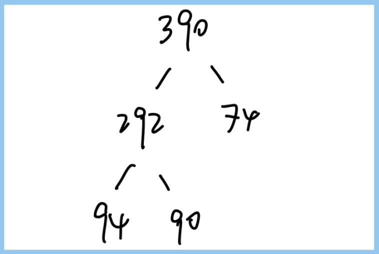
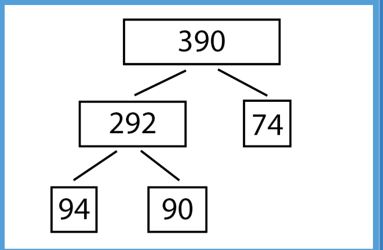
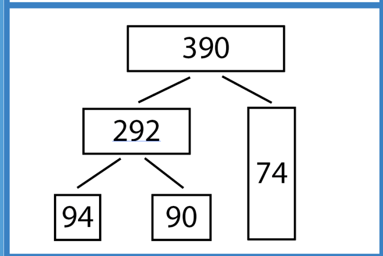
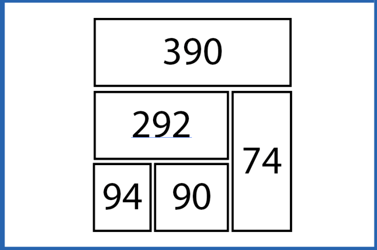
Predict One Level

Traditional Hand-written Tree

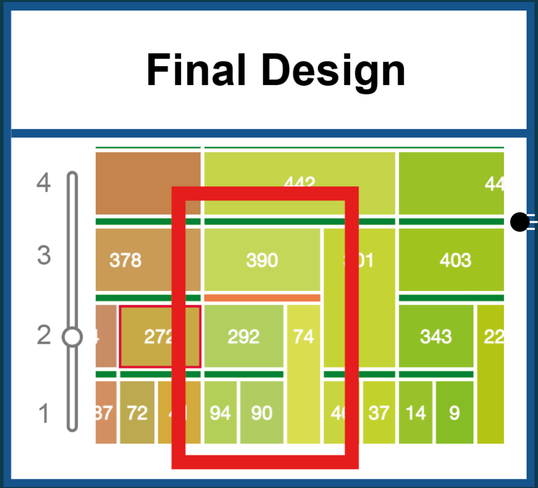


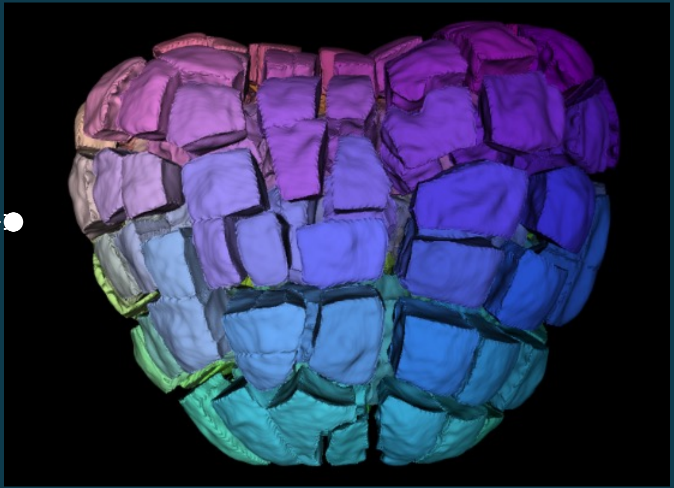


2D Abstract Hierarchy

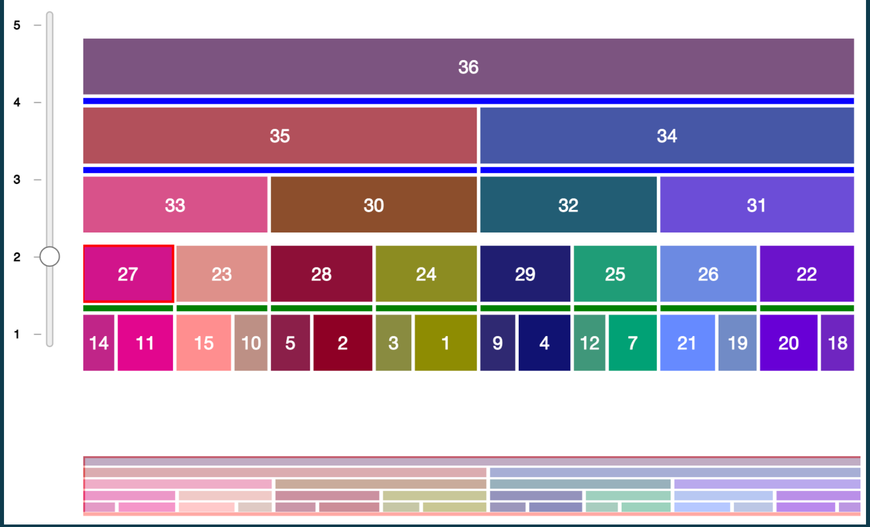
Hand writing records	Encode volume with width	Encode division time stage with height	Remove Links	Final Design
				

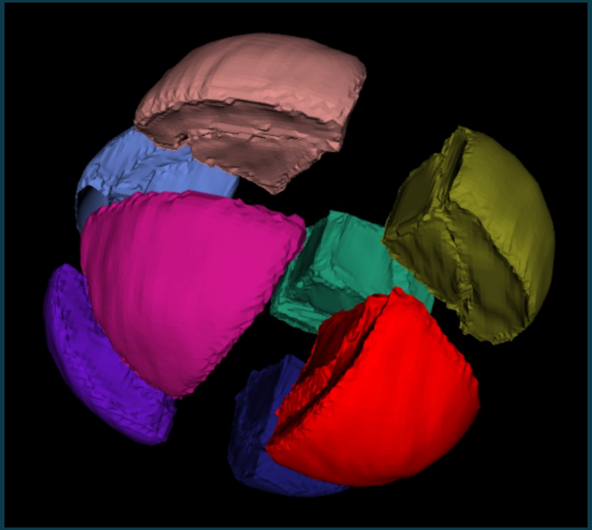
Combining 2D and 3D Representations



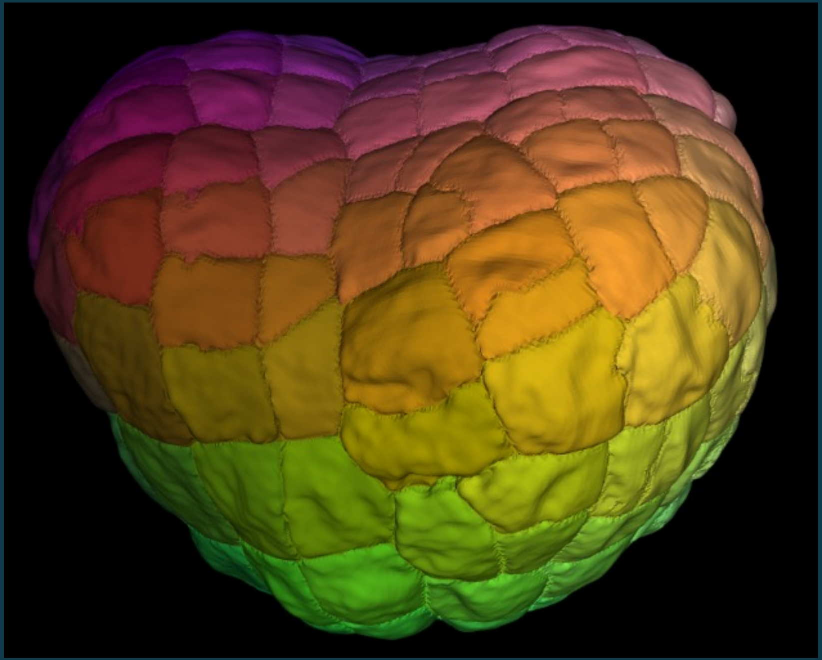


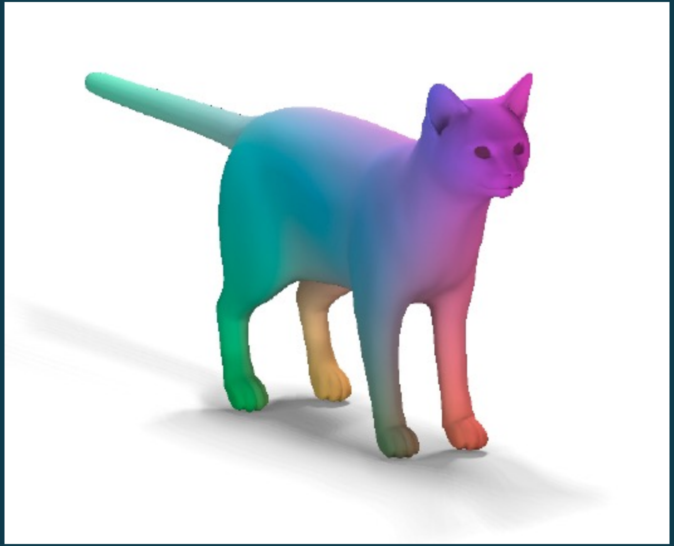
Visually and Interactively **Connected**





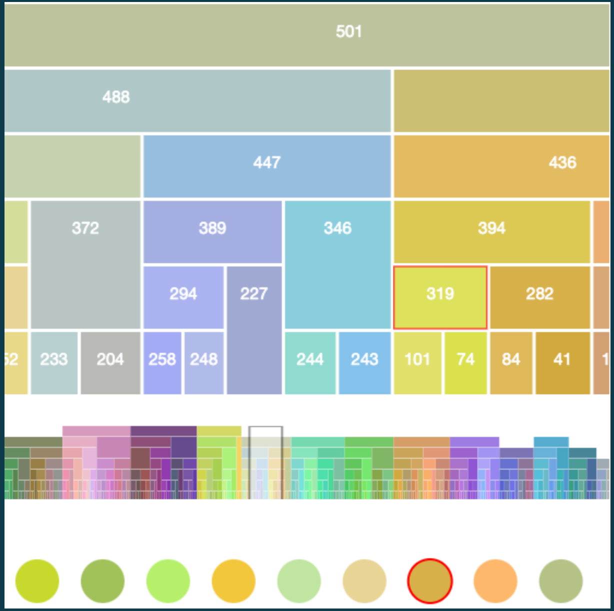
Visually and Interactively **Connected**



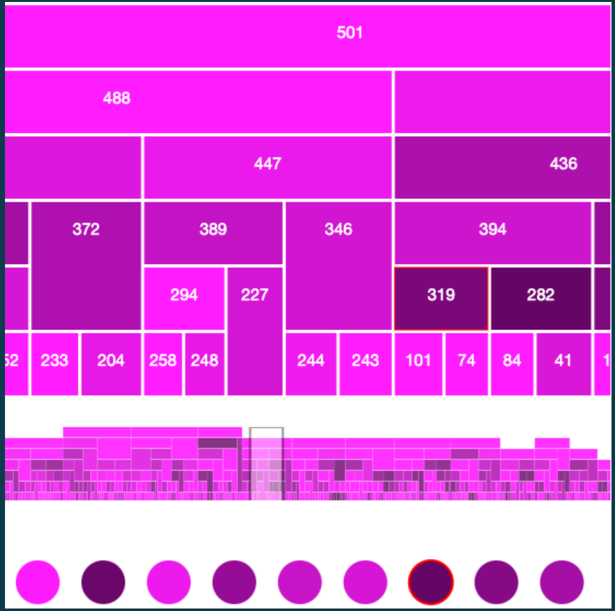


[Ovsjanikov et al. 2012]

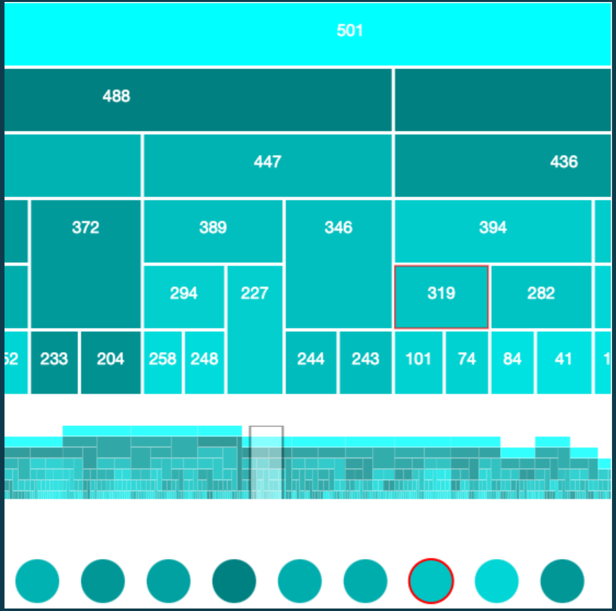
Visually and Interactively **Connected**



By 3D Positions

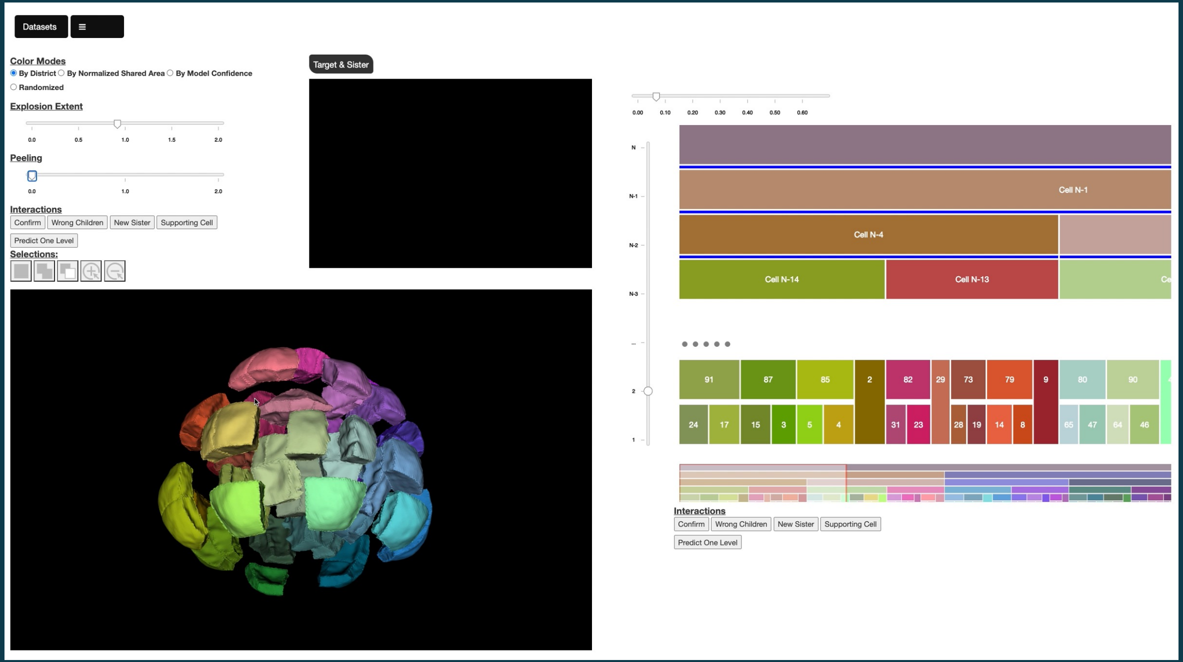


By Shared Area

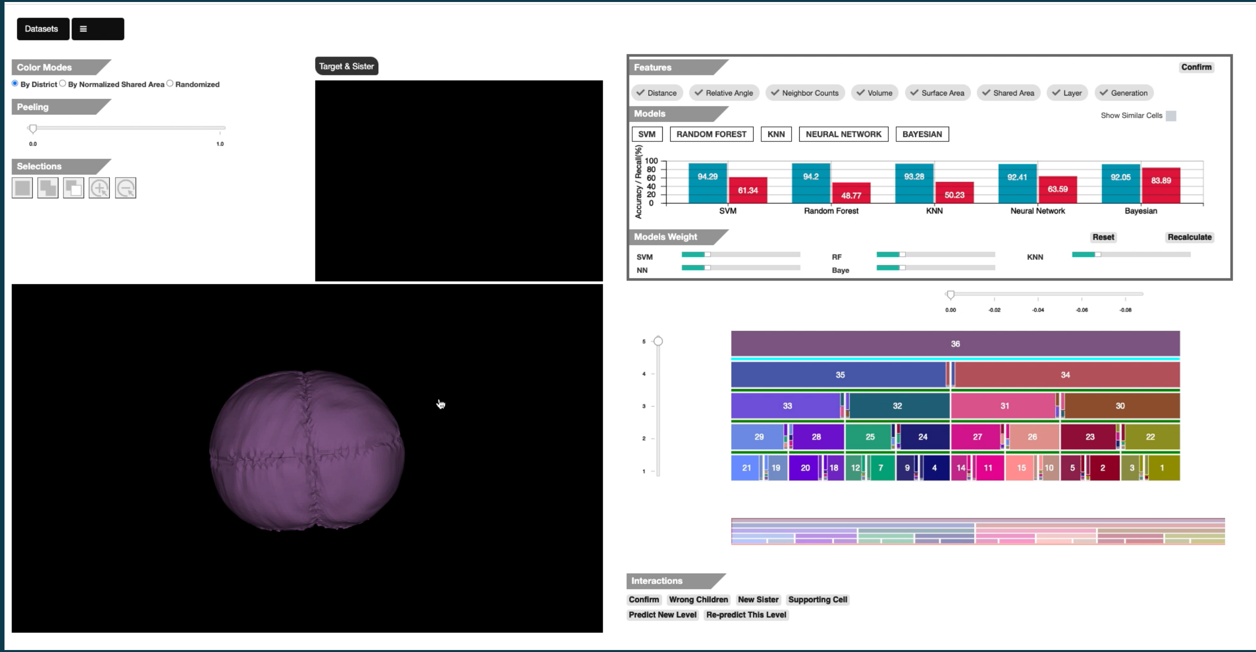


By Model Confidence

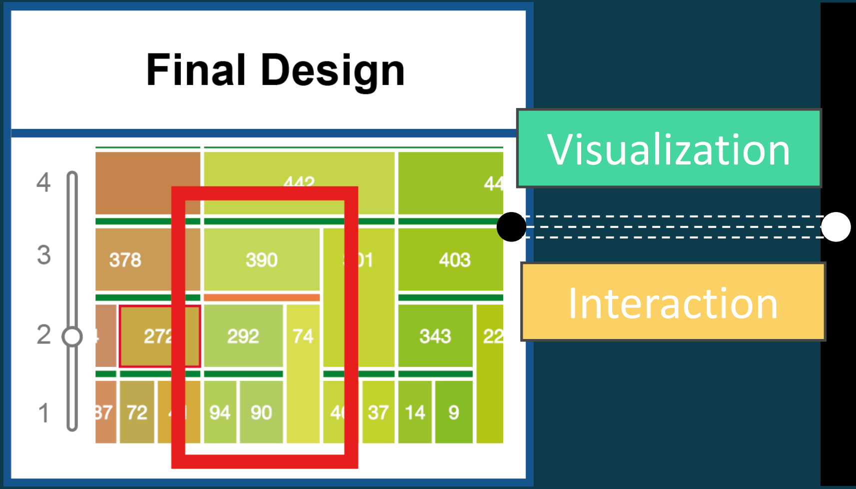
Visually and Interactively Connected

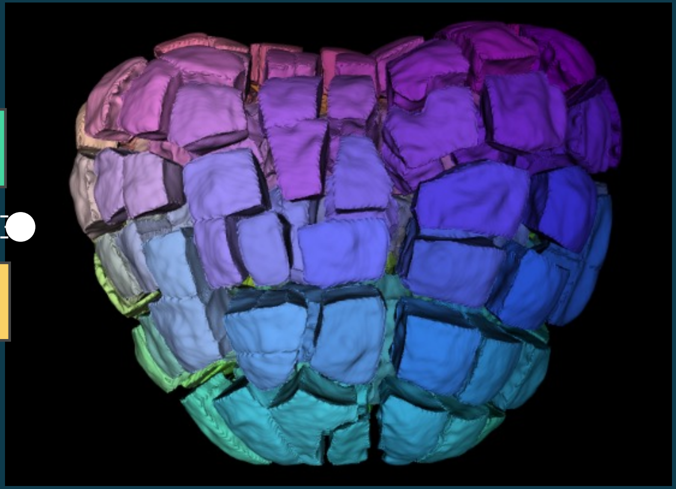


Visually and **Interactively Connected**



How to use the combination?

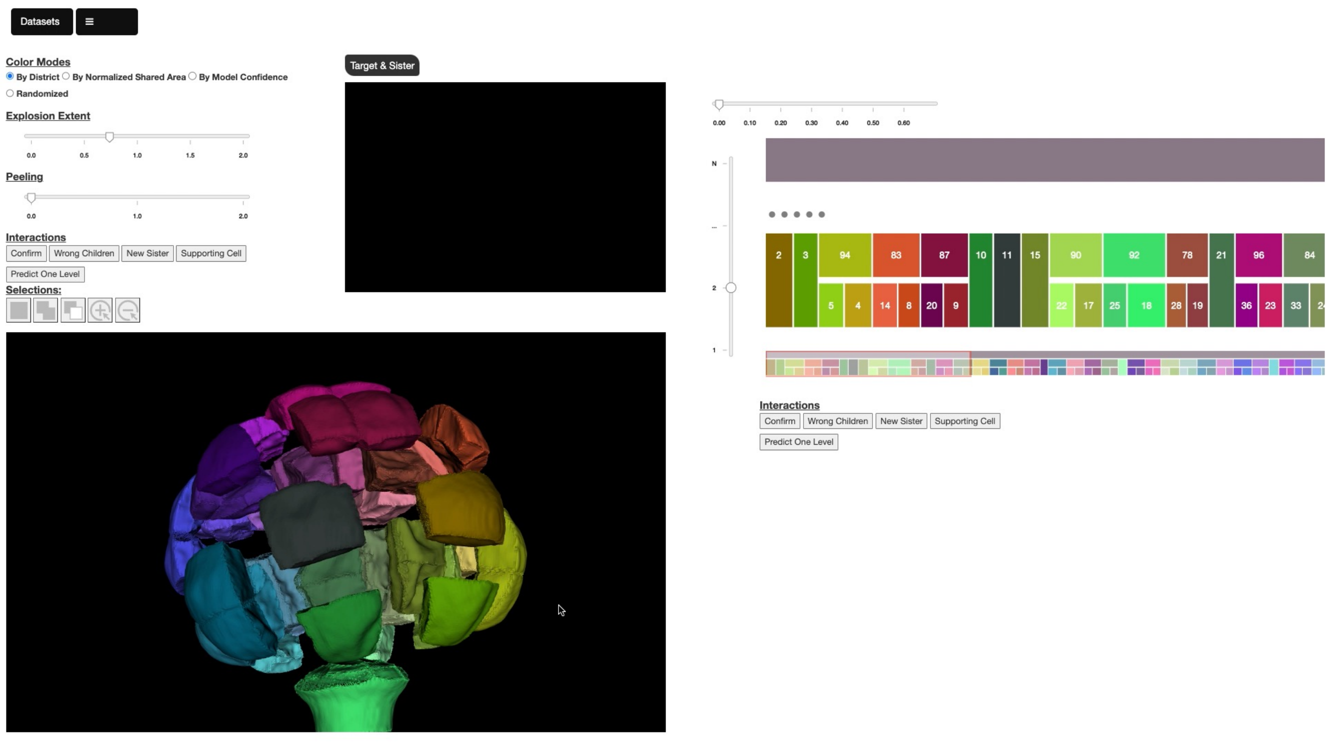




Balance the workflow

Traditional Workflow	Advanced Workflow
√ Familiar	√ Effective
X Less effective	X Learning costs

Traditional Manual Assignments



Advanced Workflow

Two-direction of Hierarchy Building

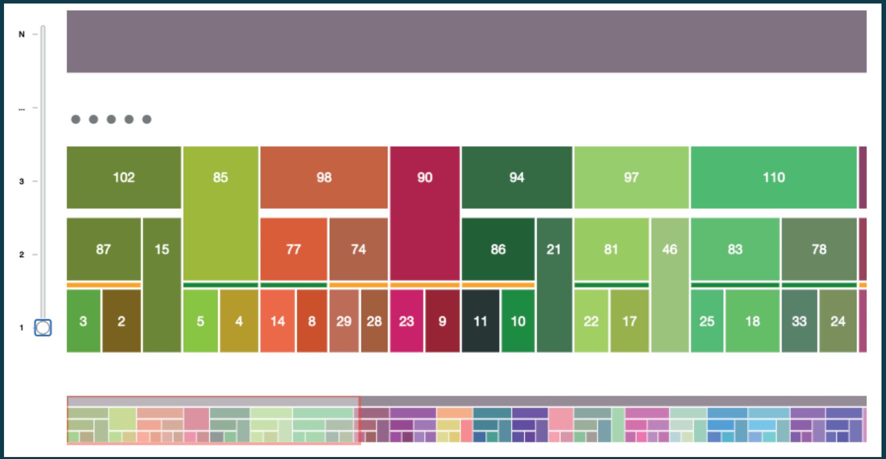
Enable biologists to build the hierarchy tree from both top-down and bottom-up approaches.

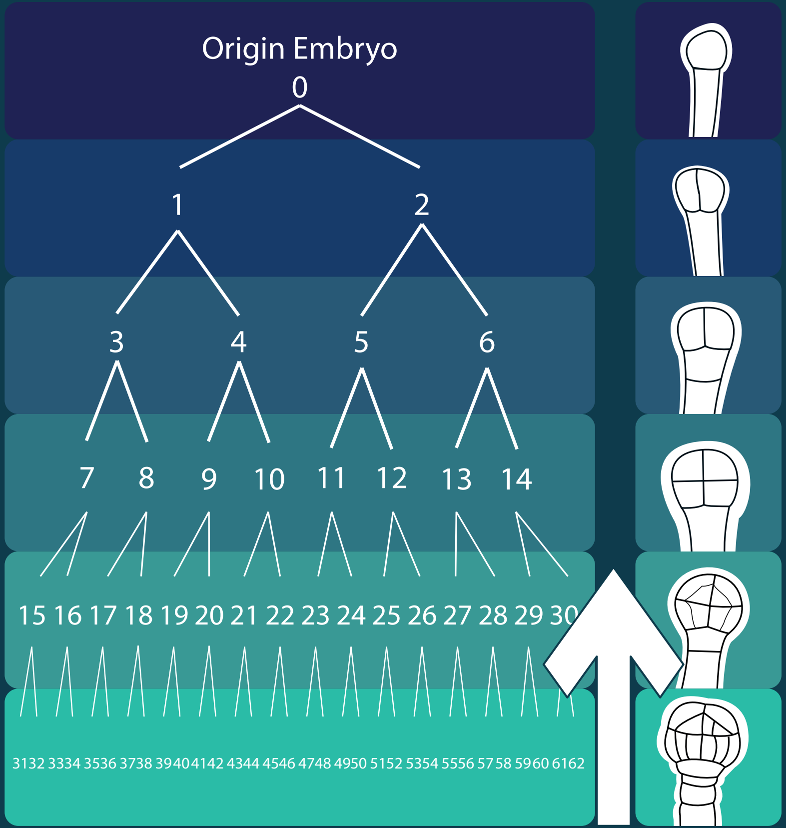
Machine Learning Predictions

Use ML to predict a single level as a basis for biologists to check and correct.

Bottom-up Approach

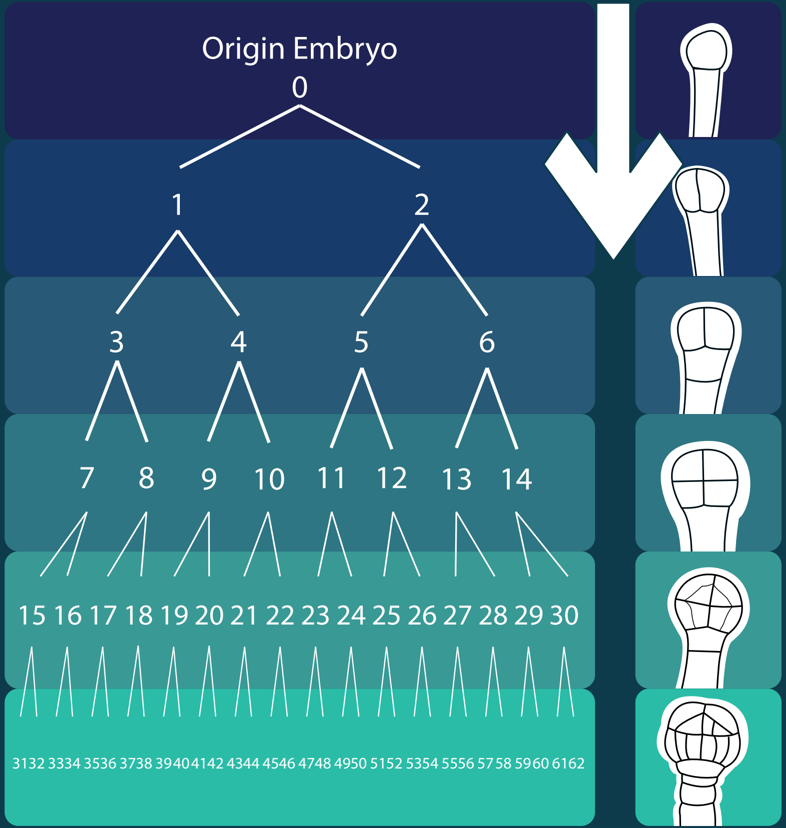
The traditional way to build the hierarchy.

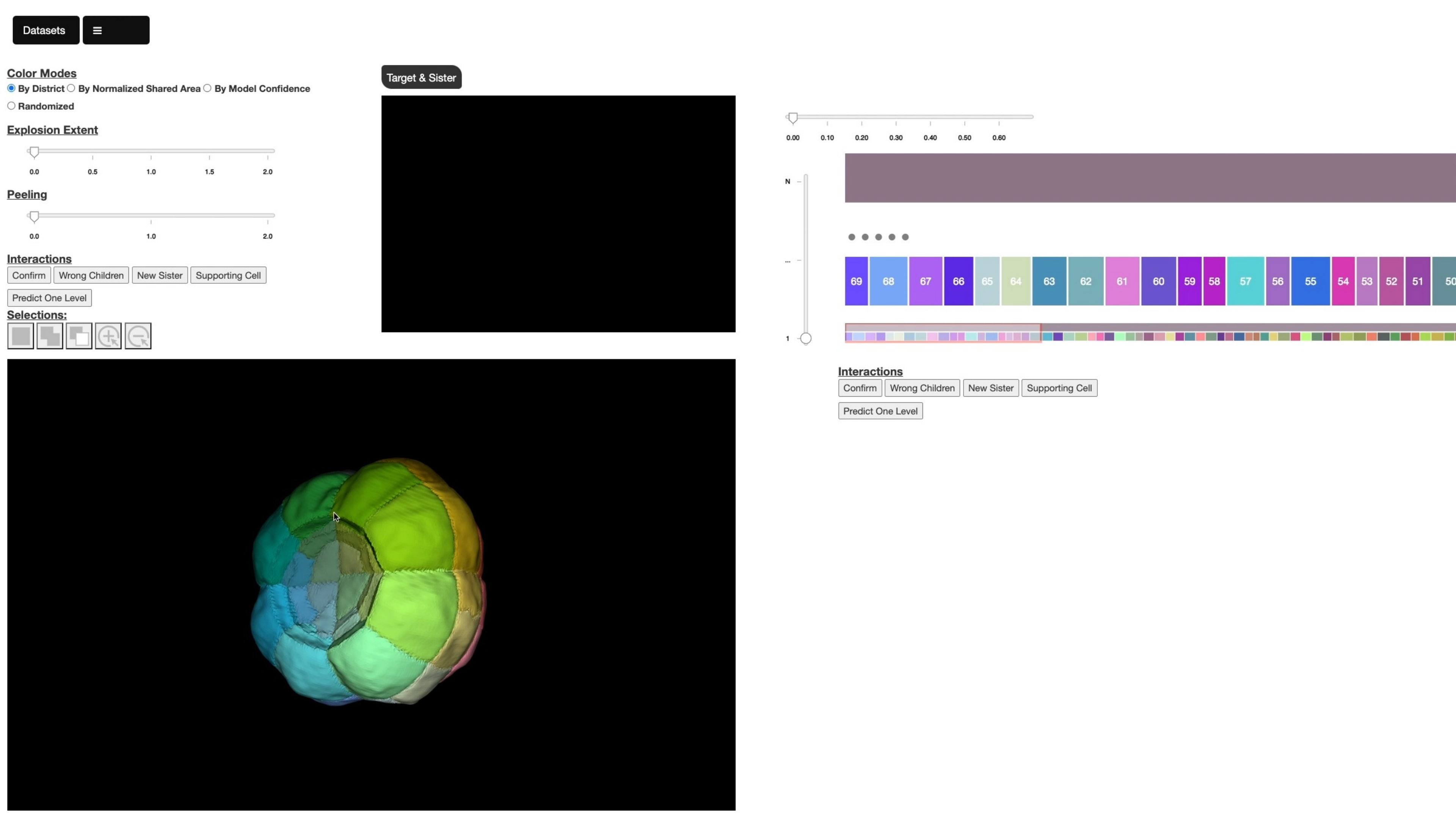




Top-down Approach

Biologists have ideas about how the embryo could be divided in the beginning.



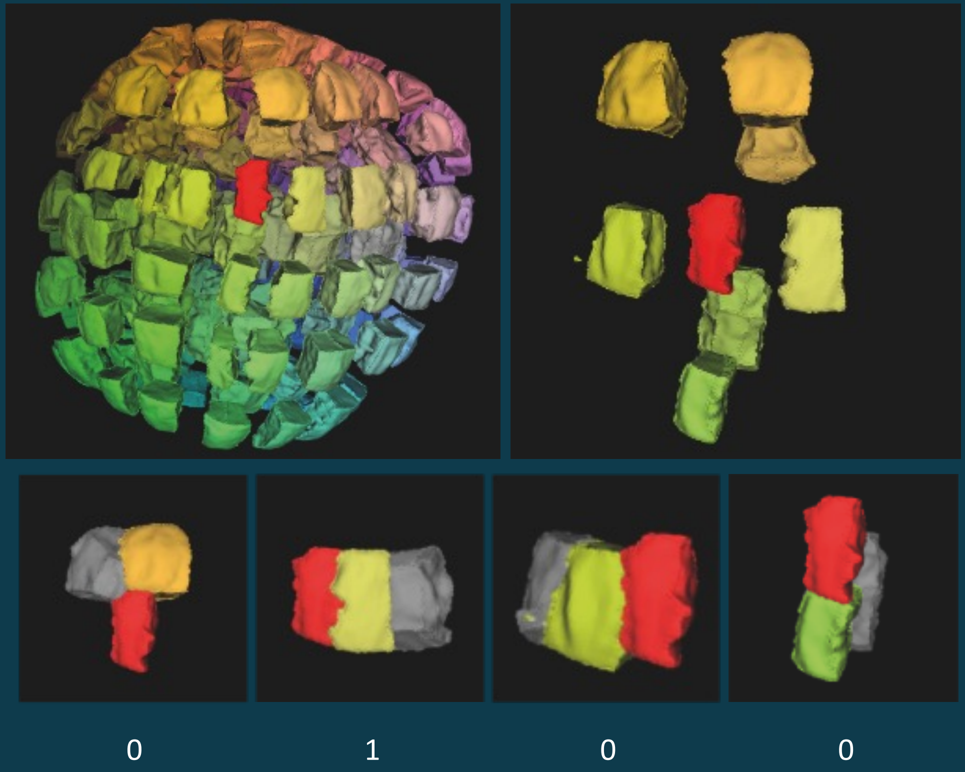

Datasets
Color Modes
By District
By Normalized Shared Area
By Model Confidence
Randomized
Explosion Extent
0.0 0.5 1.0 1.5 2.0
Peeling
0.0 1.0 2.0
Interactions
Confirm
Wrong Children
New Sister
Supporting Cell
Predict One Level
Selections:
Target & Sister
0.00 0.10 0.20 0.30 0.40 0.50 0.60
N
...
1
69 68 67 66 65 64 63 62 61 60 59 58 57 56 55 54 53 52 51 50
Interactions
Confirm
Wrong Children
New Sister
Supporting Cell
Predict One Level

Machine Learning Model

Binary Classification Problem:

Sisters (1) / Non-sisters (0)

- Sequential Neural Network
- 93 Embryo datasets
- 47132 pairs
- 12 features



Machine Learning Model

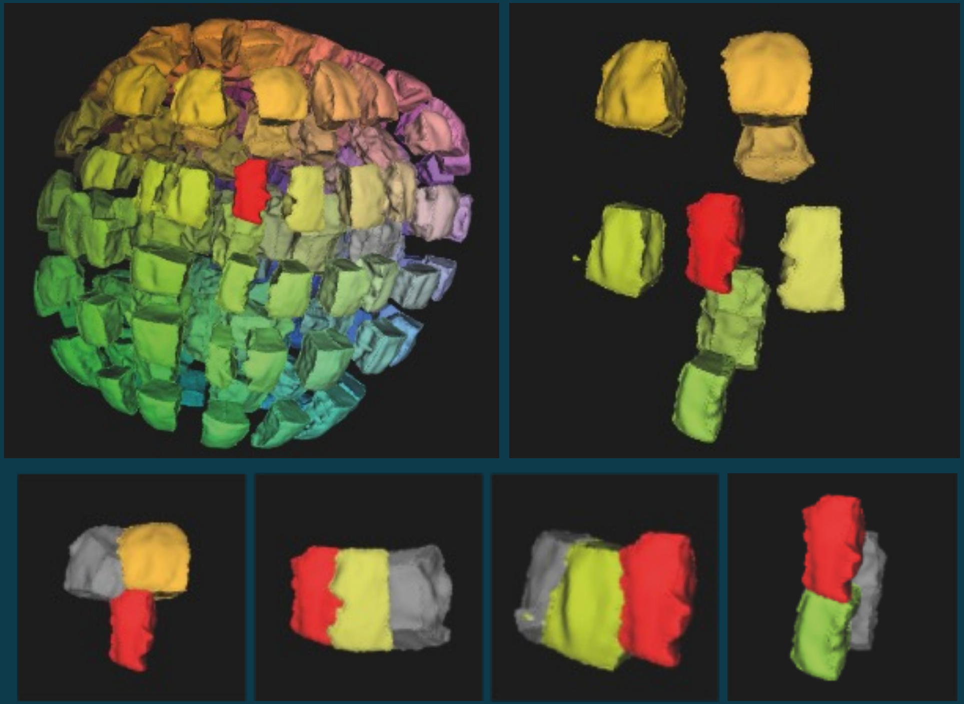
Binary Classification Problem:

Sisters (1) / Non-sisters (0)

- Sequential Neural Network

! 93 Embryo datasets

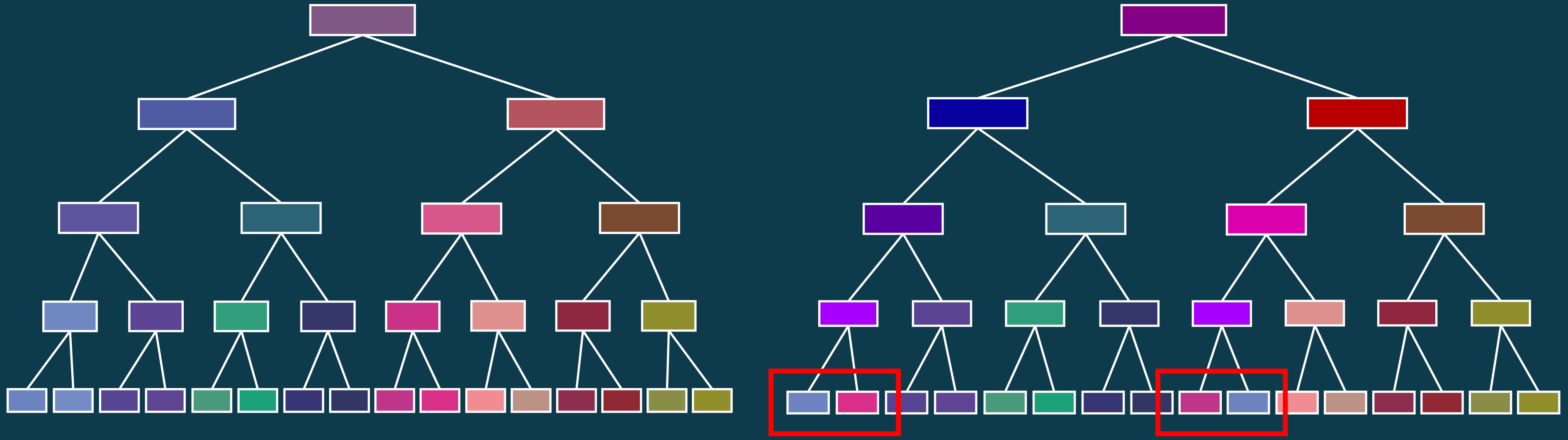
- 47132 pairs
- 12 features



0 1 0 0

Machine Learning Model

Limited training datasets

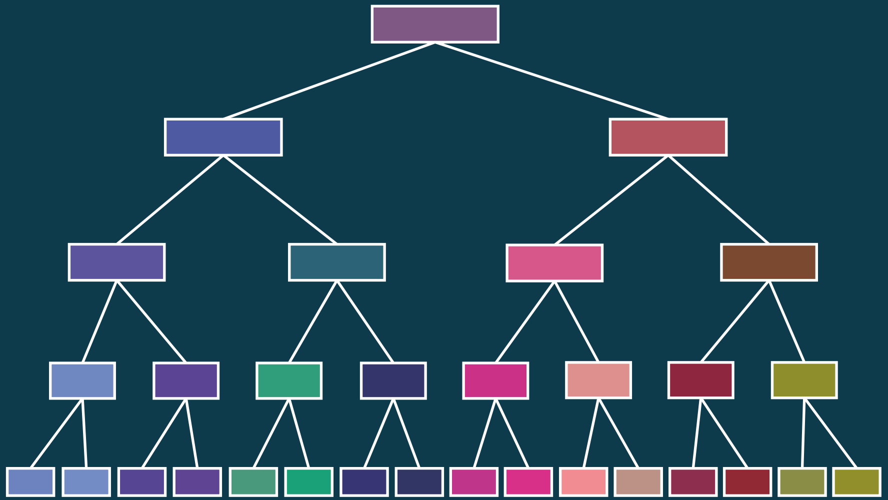


Machine Learning Model

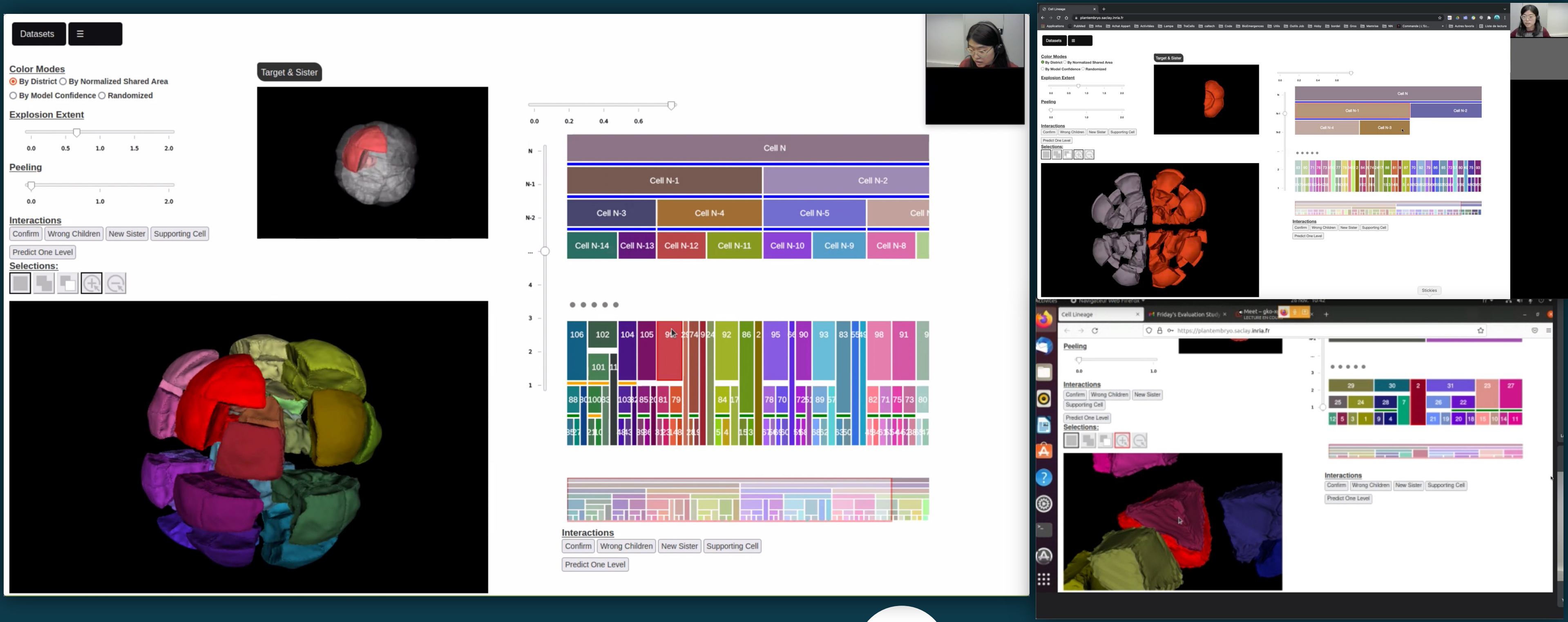
Limited training datasets

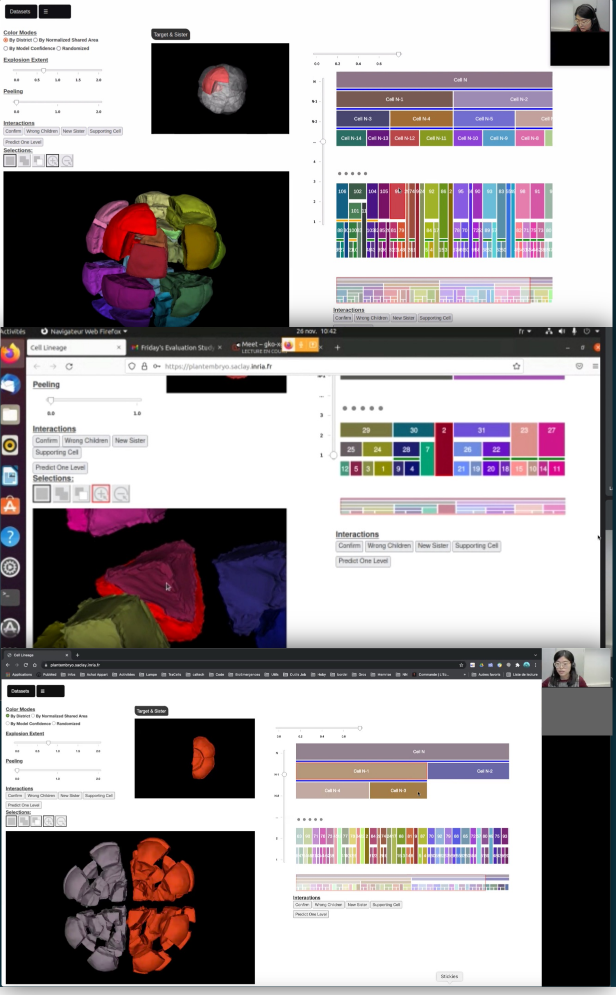
Machine Learning Model

Limited training datasets



Evaluation Study





Evaluation Study

- Biologists took time to learn the functionalities of LineageD.
- They used both the 2D hierarchy and 3D views.
- They valued the visual representations in 2D and 3D.
- They thought LineageD helps to understand the embryos' development better.

Publication

Jiayi Hong, Alain Trubuil, and Tobias Isenberg. “LineageD: An Interactive Visual System for Plant Cell Lineage Assignments based on Correctable Machine Learning”. In: Computer Graphics Forum 41.3 (June 2022), pp. 195–207.

DOI: 10.1111/cgf.14533

Eurographics Conference on Visualization (EuroVis) 2022
R. Borgo, G. E. Marai, and T. Schreck
(Guest Editors)

Volume 41 (2022), Number 3

LineageD: An Interactive Visual System for Plant Cell Lineage Assignments based on Correctable Machine Learning

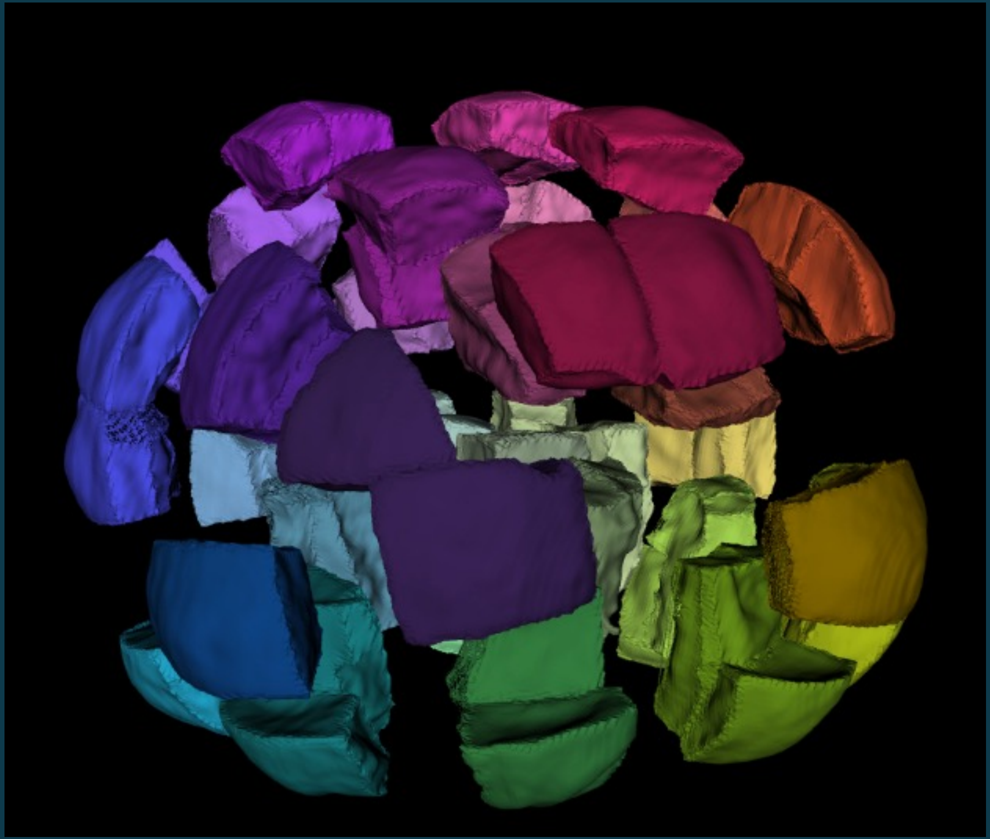
Jiayi Hong[1], Alain Trubuil[2], and Tobias Isenberg[1]

[1]Université Paris-Saclay, CNRS, Inria, France
[2] Université Paris-Saclay, InraE, France

Abstract

We describe LineageD—a hybrid web-based system to predict, visualize, and interactively adjust plant embryo cell lineages. Currently, plant biologists explore the development of an embryo and its hierarchical cell lineage manually, based on a 3D dataset that represents the embryo status at one point in time. This human decision-making process, however, is time-consuming, tedious, and error-prone due to the lack of integrated graphical support for specifying the cell lineage. To fill this gap, we developed a new system to support the biologists in their tasks using an interactive combination of 3D visualization, abstract data visualization, and correctable machine learning to modify the proposed cell lineage. We use existing manually established cell lineages to obtain a neural network model. We then allow biologists to use this model to repeatedly predict assignments of a single cell division stage. After each hierarchy level prediction, we allow them to interactively adjust the machine learning based assignment, which we then integrate into the pool of verified assignments for further predictions. In addition to building the hierarchy this way in a bottom-up fashion, we also offer users to divide the whole embryo and create the hierarchy tree in a top-down fashion for a few steps, improving the ML-based assignments by reducing the potential for wrong predictions. We visualize the continuously updated embryo and its hierarchical development using both 3D spatial and abstract tree representations, together with information about the model's confidence and spatial properties. We conducted case study validations with five expert biologists to explore the utility of our approach and to assess the potential for LineageD to be used in their daily workflow. We found that the visualizations of both 3D representations and abstract representations help with decision making and the hierarchy tree top-down building approach can reduce assignments errors in real practice.

CCS Concepts
• ***Human-centered computing*** → ***Scientific visualization; User interface toolkits;***



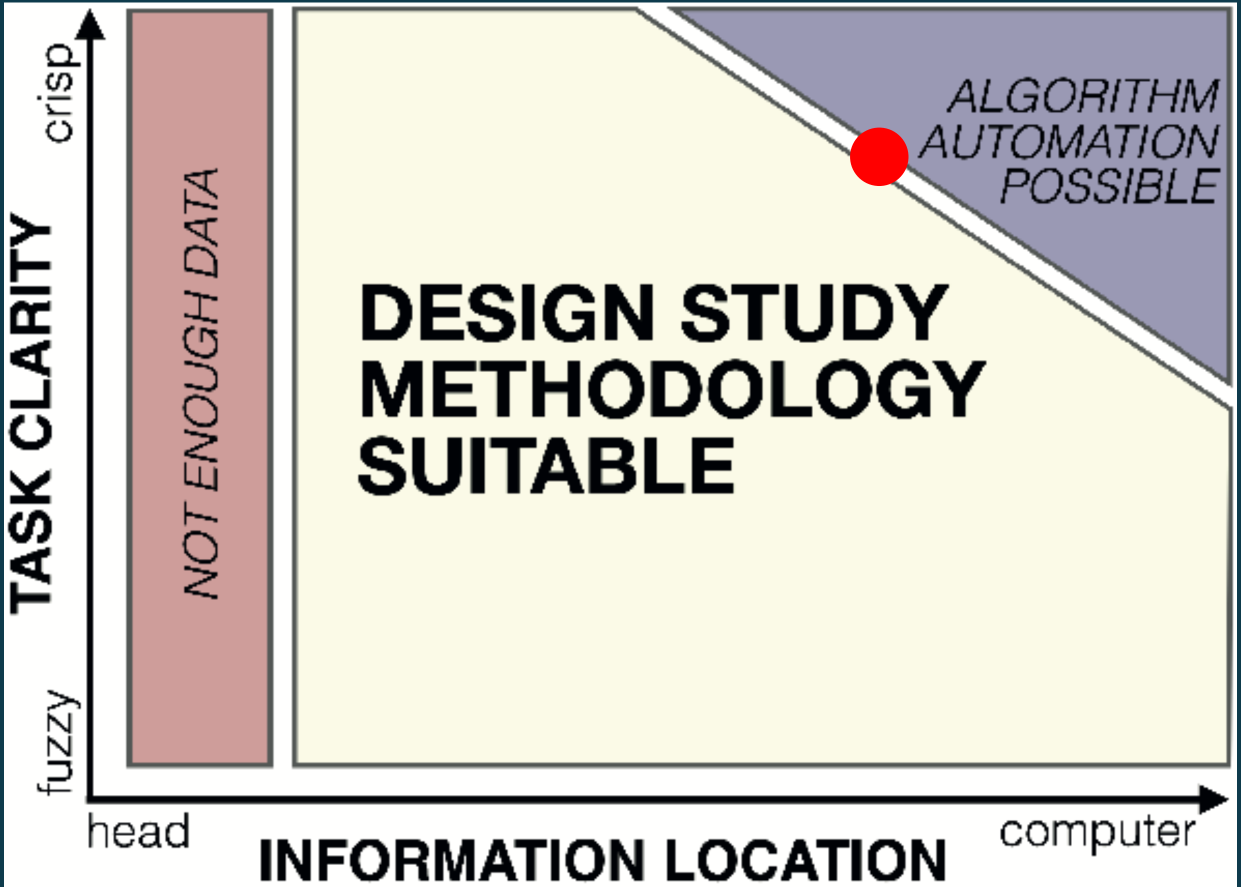
ML Results not satisfying

How could we help biologists to have a better collaboration with ML in the case where **the training datasets are limited**?

04

Third Project

LineageD+



[Sedlmair et al. 2012]

Limited Training Datasets

Create the “best” model

Enable users to make choices

Our Solution – Human-AI Teaming

Make use of machine learning

We need to try our best to improve ML performances.

Allow people to control the results

Human beings should have full control over the final decisions.

Our Solution – Human-AI Teaming

Make use of machine learning

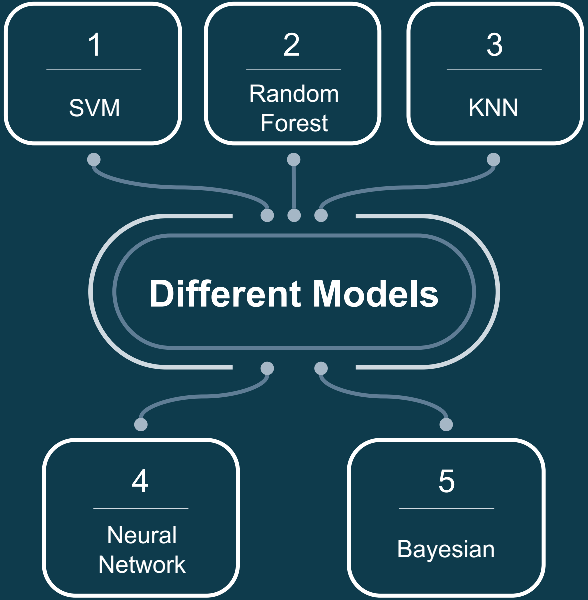
We need to try our best to improve ML performances.

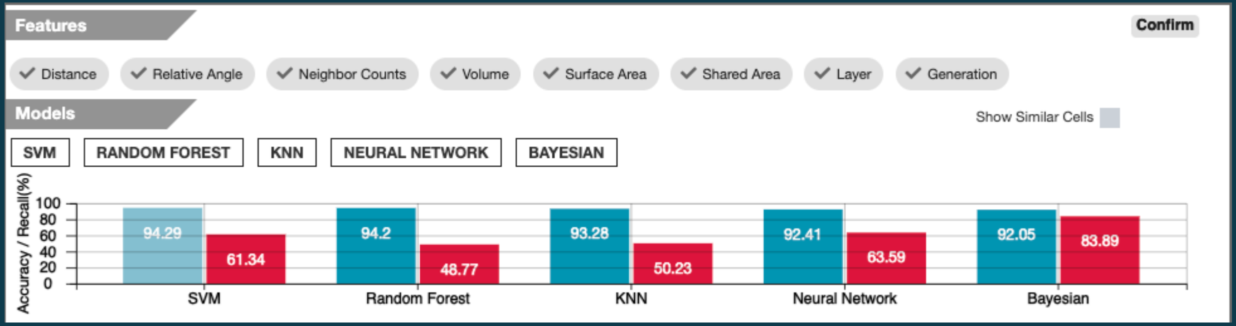
Model Training

Different Models

1. SVM
2. Random Forest
3. KNN
4. Neural Network
5. Bayesian

Model Training

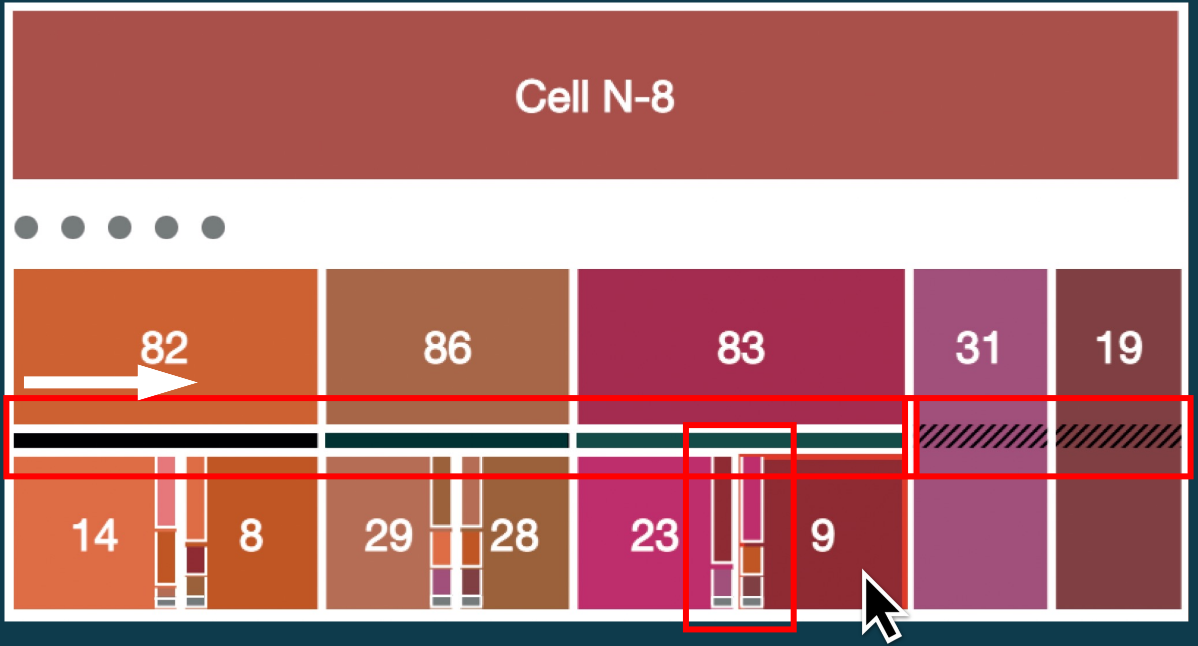




Accuracy

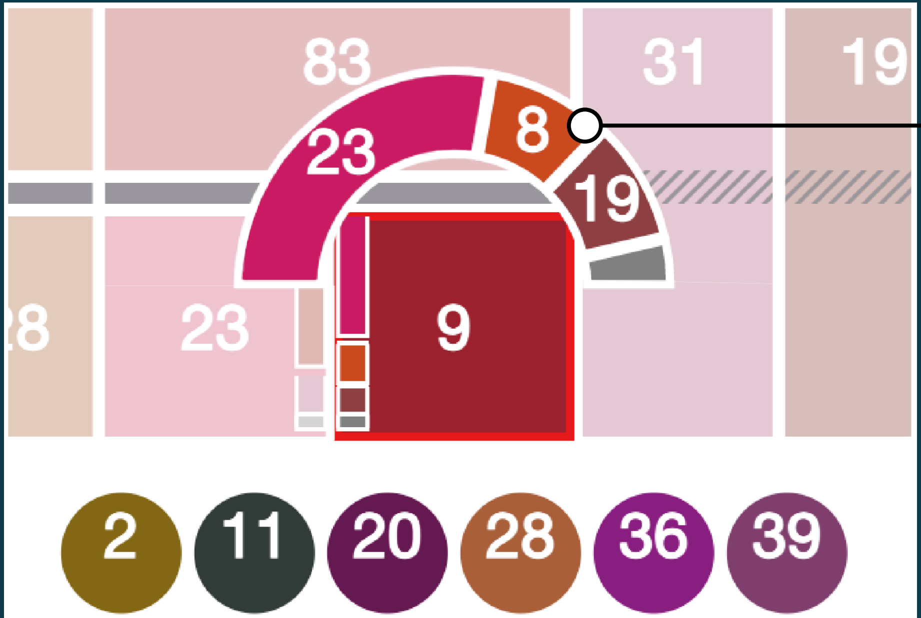
Recall

Prediction Visualization



Combining the prediction results from five models, we visualized them with stacked bar charts on each node.

Prediction Visualization

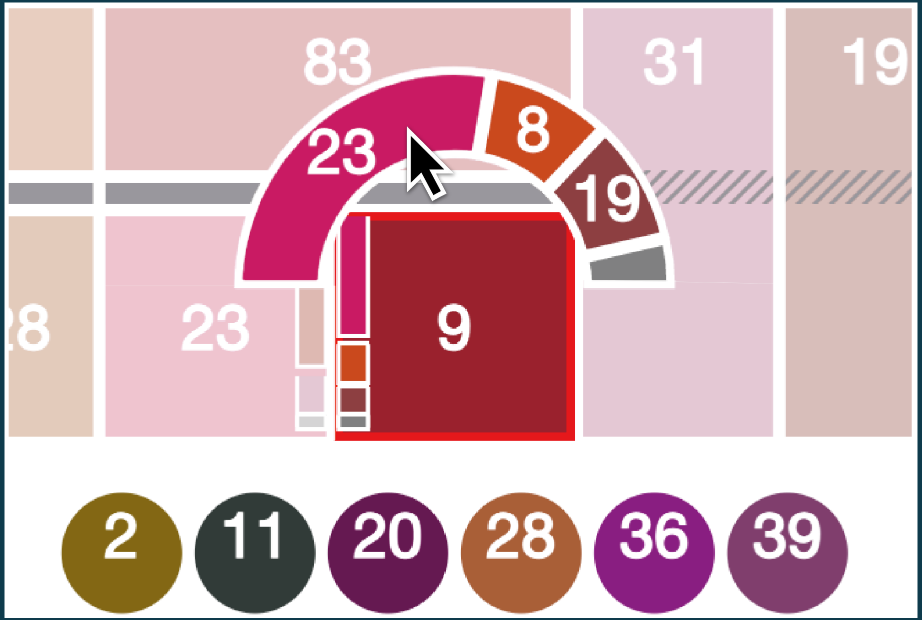


All predicted sisters

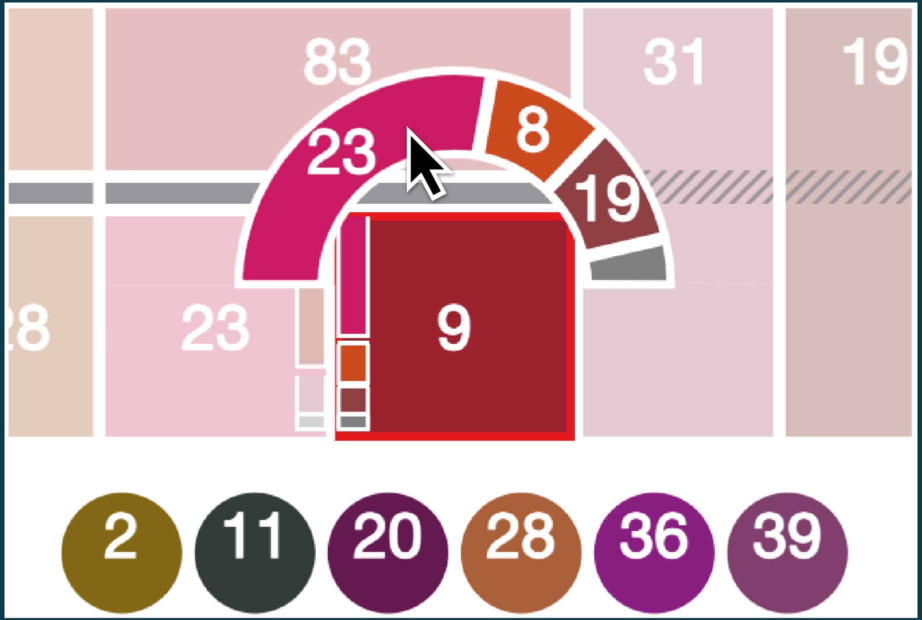
Vertical thumbnail of all predictions

All the other neighboring cells

Prediction Visualization

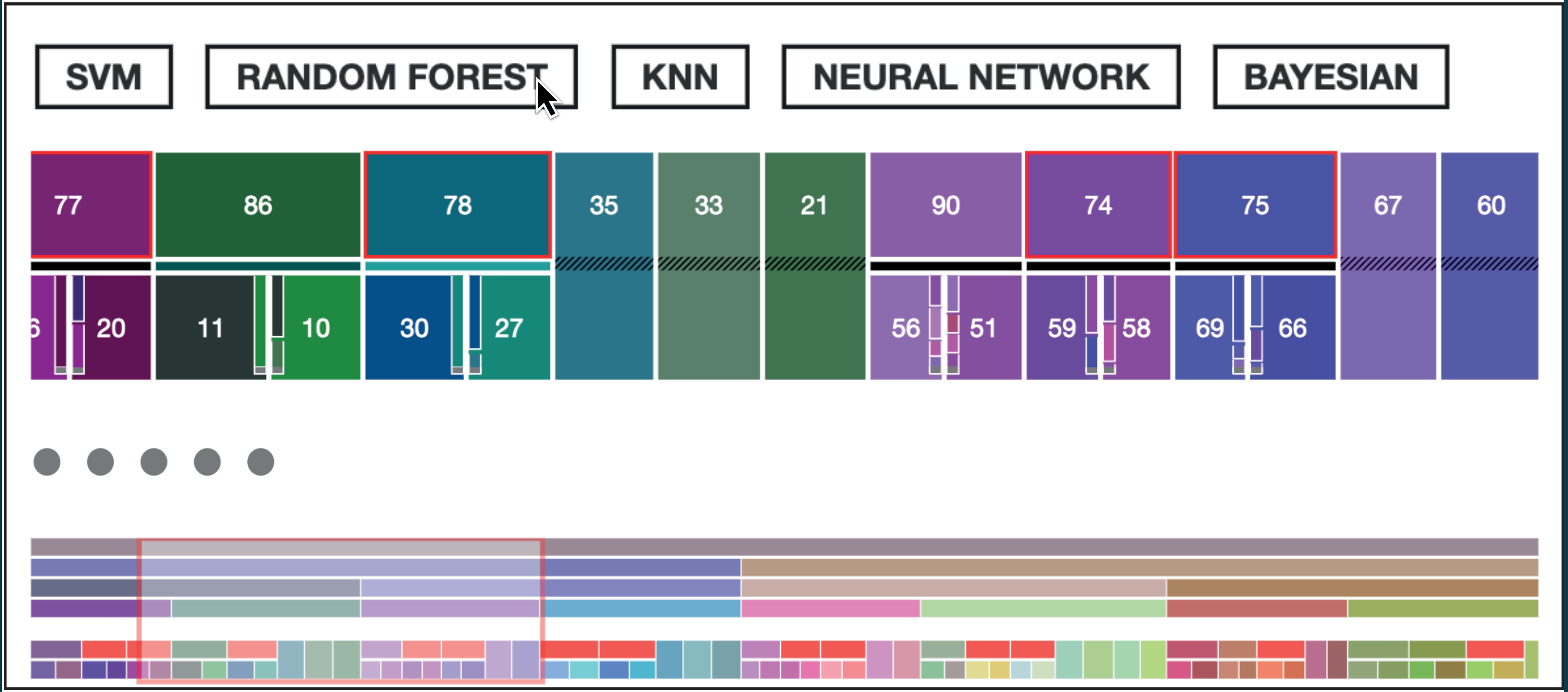


Prediction Visualization



neuralNetwork	23
knn	
bayesian	
svm	23
randomForest	23

Individual Model Prediction



Our Solution – Human-AI Teaming

Make use of machine learning

We need to try our best to improve ML performances.

Allow people to control the results

Human beings should have full control over the final decisions.

Our Solution – Human-AI Teaming

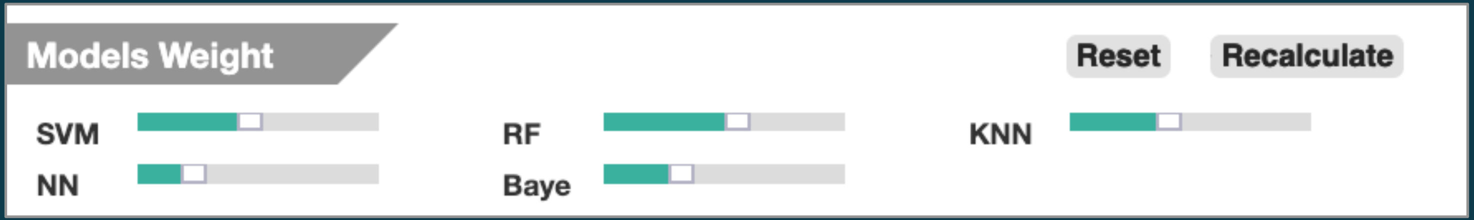
Allow people to control the results

Human beings should have full control over the final decisions.

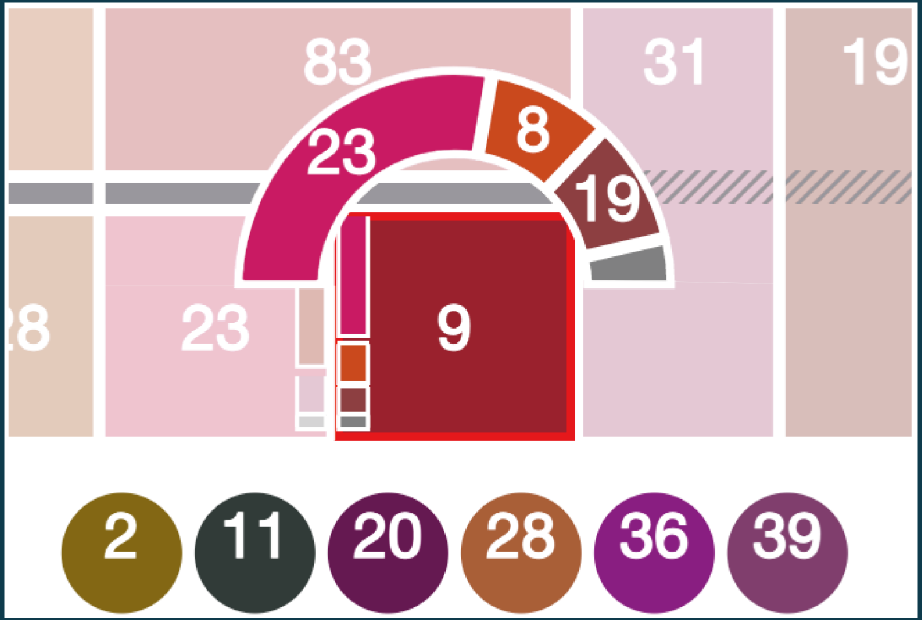
The overall predictions for each cell

Models Weight

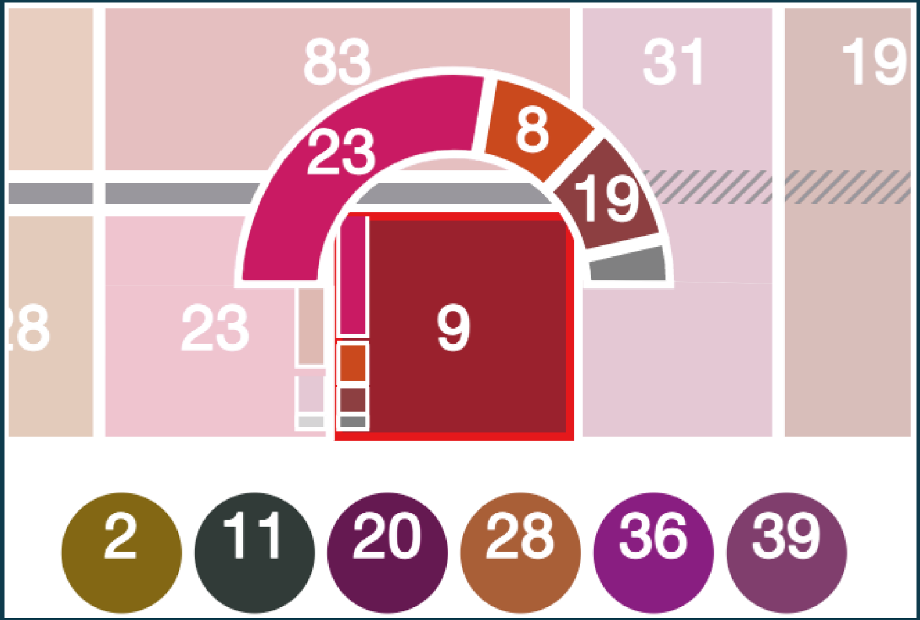
The visualization of model prediction confidence



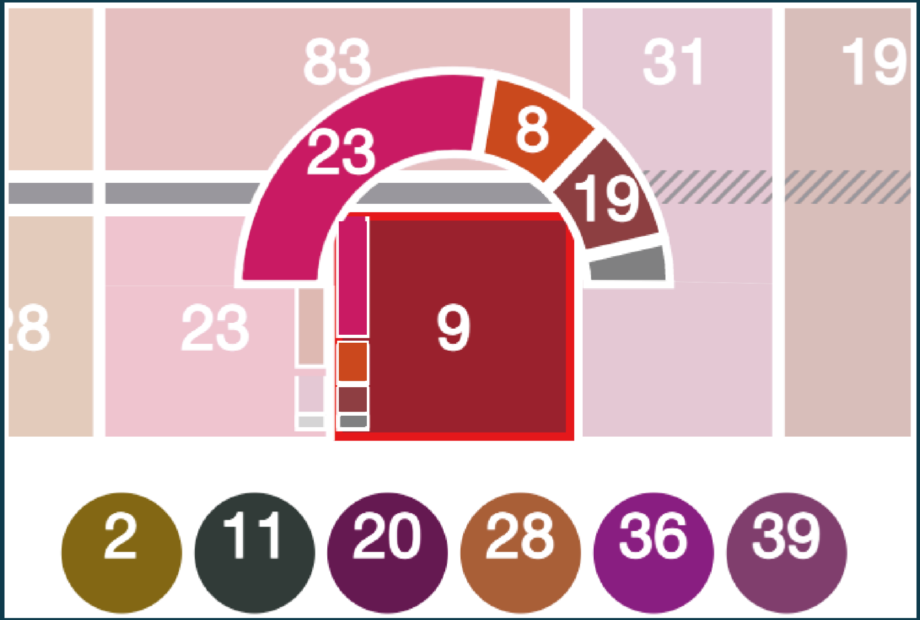
Prediction Visualization



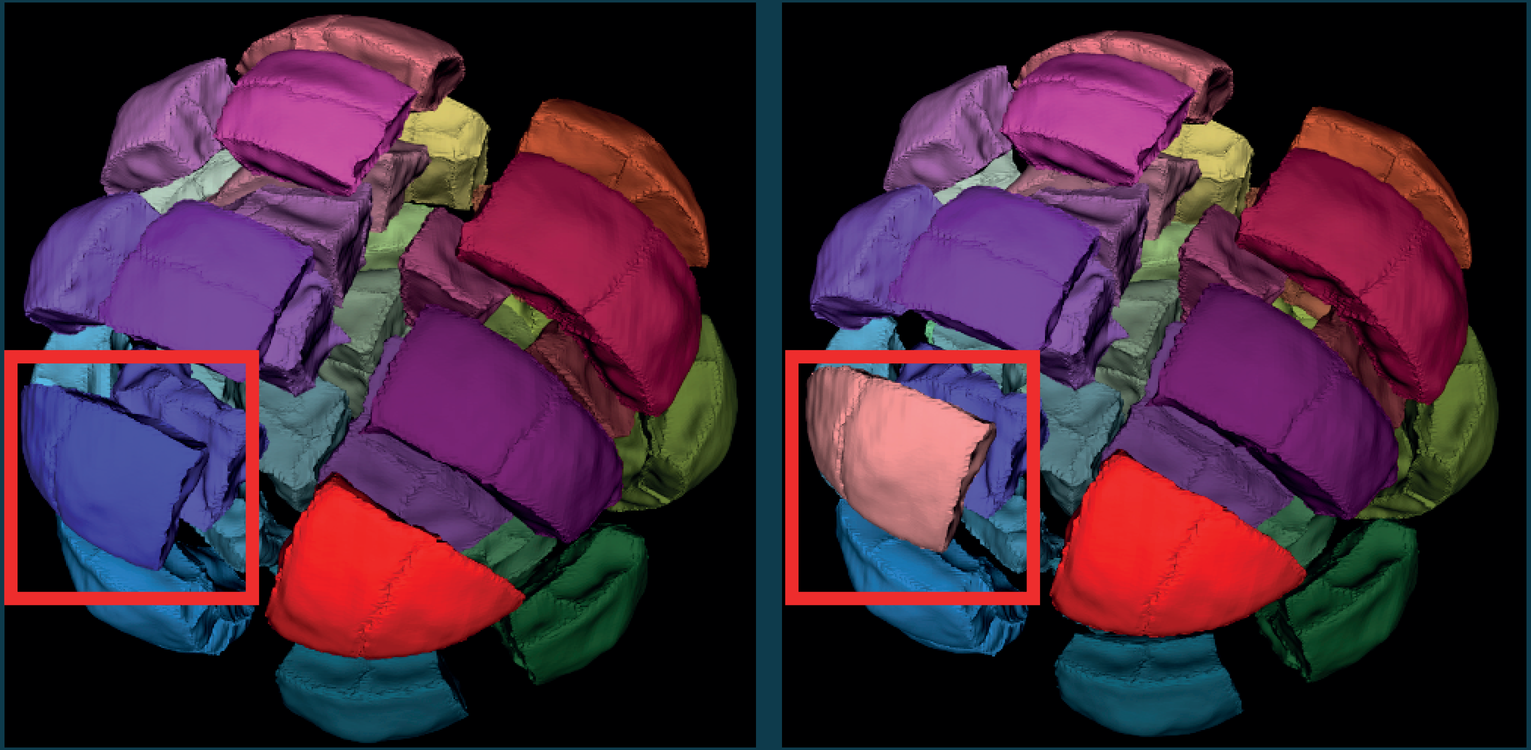
Prediction Visualization



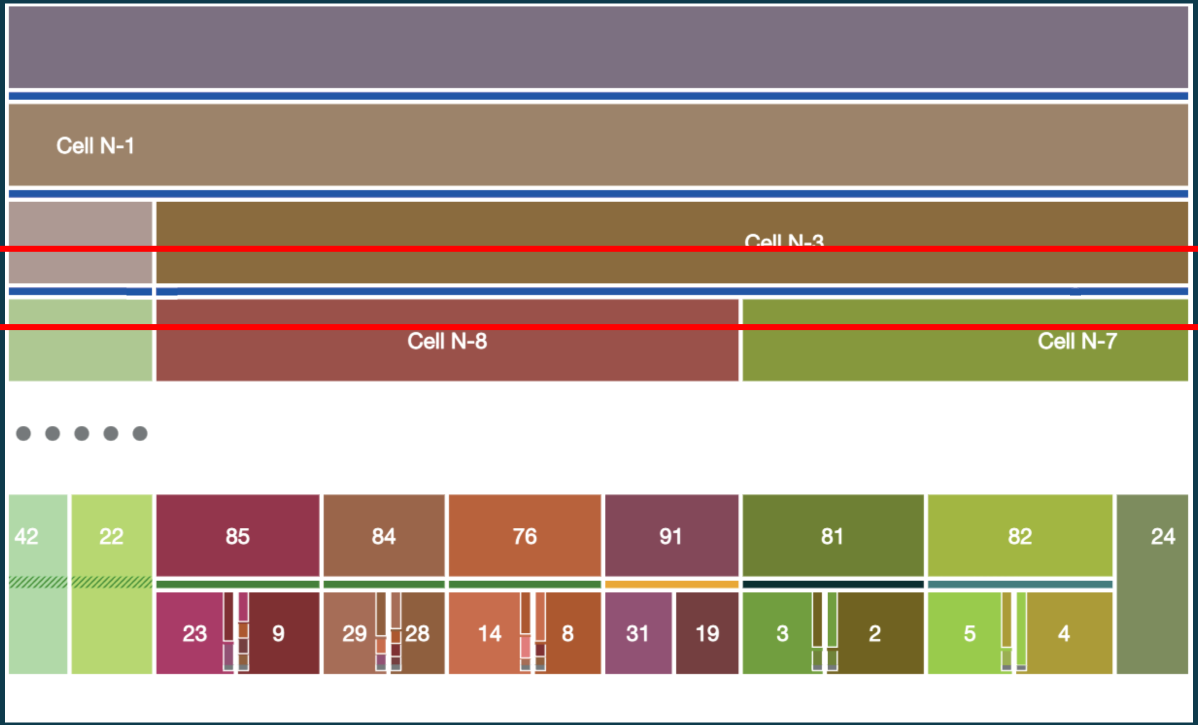
Prediction Visualization



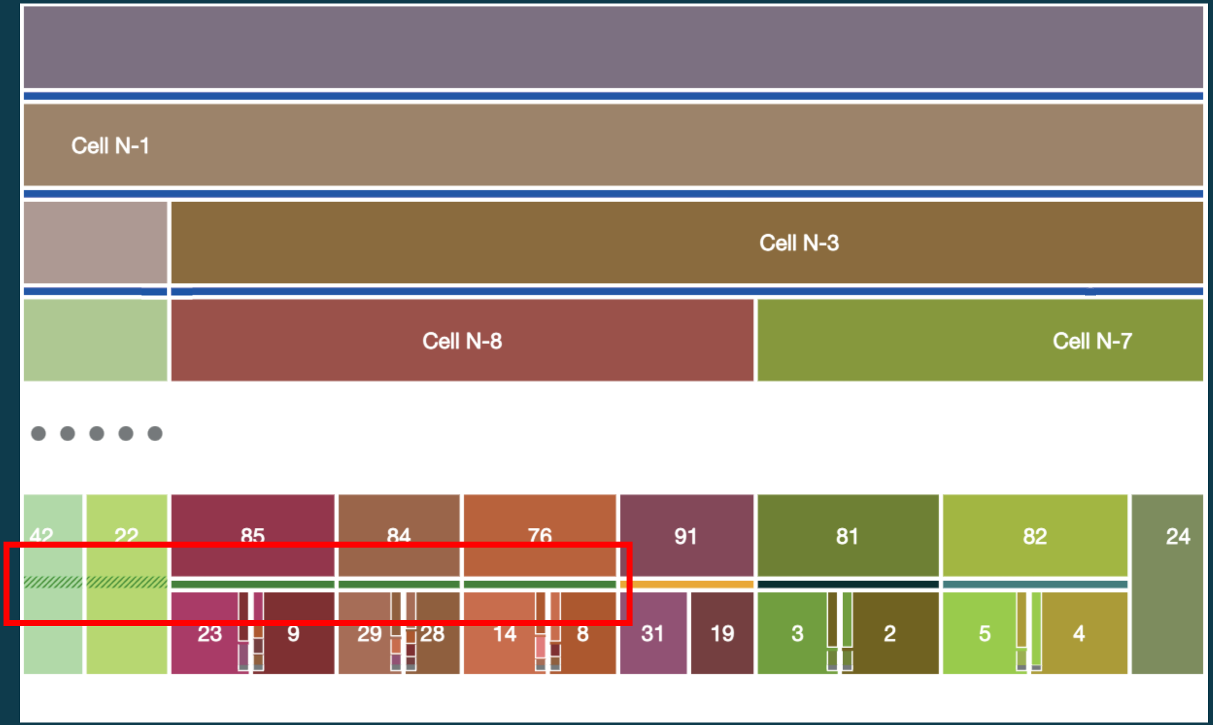
Similar Patterns



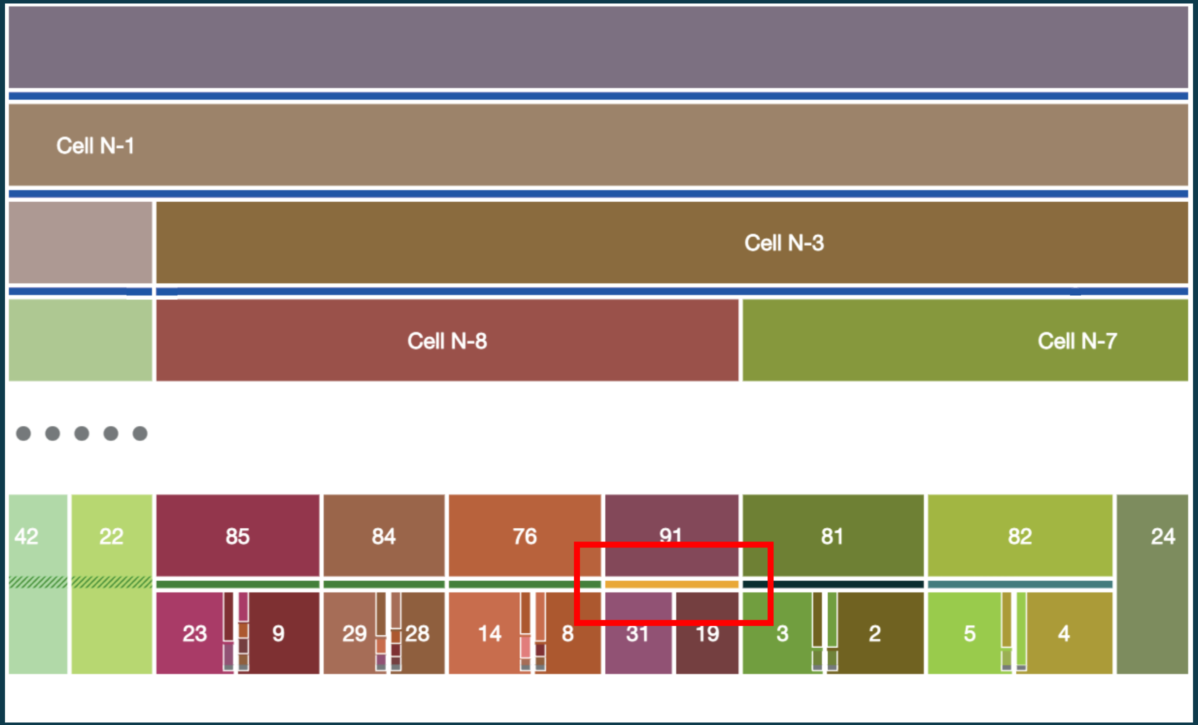
Differentiate Manual Assignments



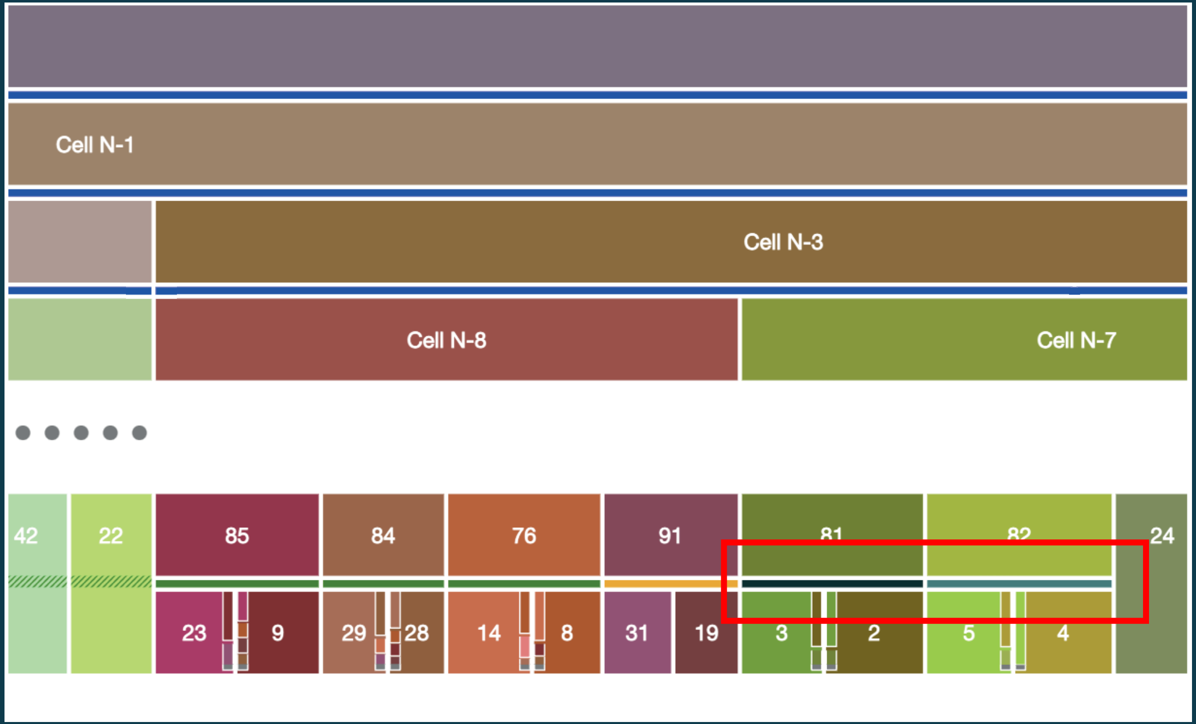
Differentiate Manual Assignments

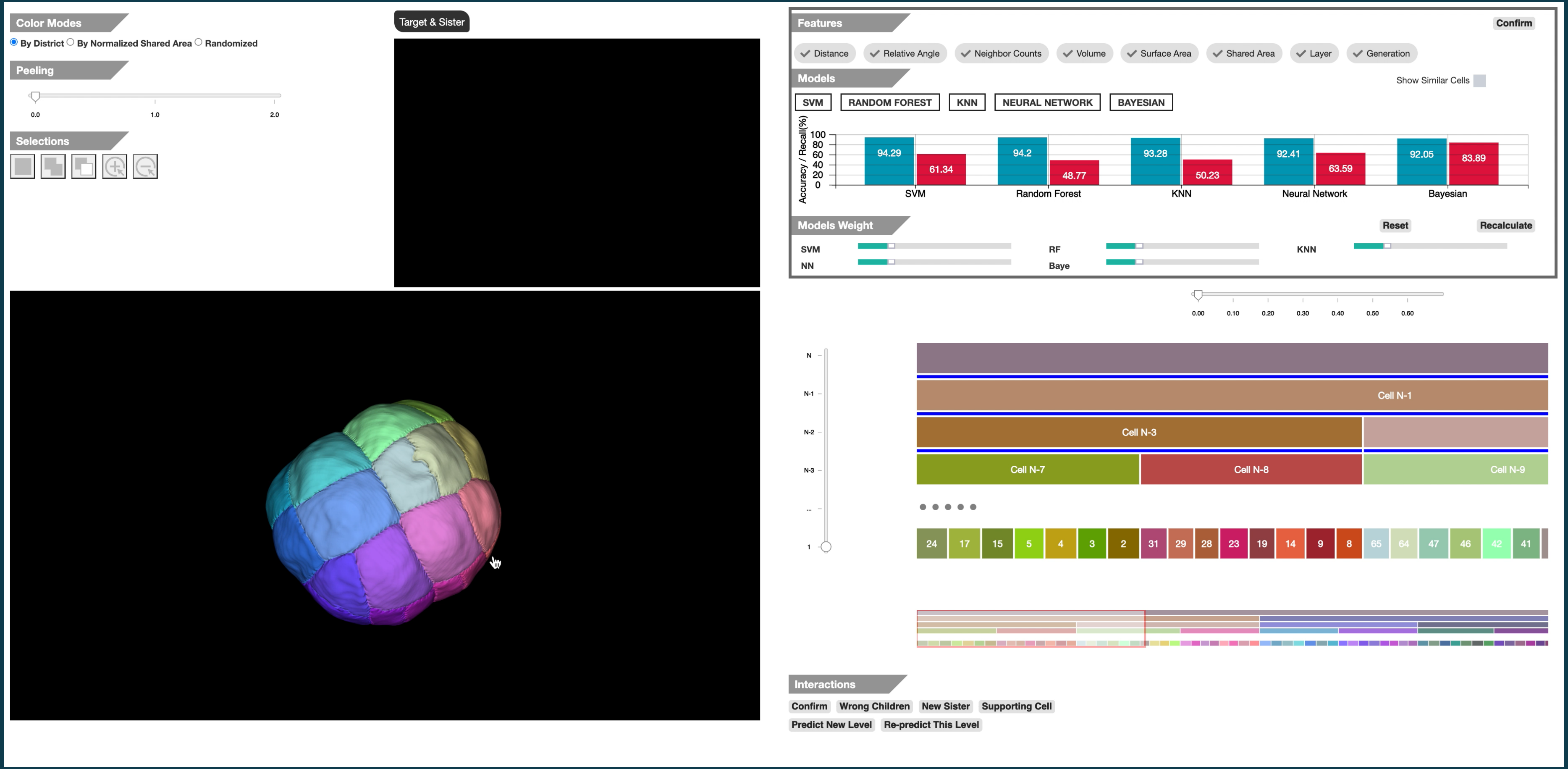


Differentiate Manual Assignments



Differentiate Manual Assignments




Color Modes
By District
By Normalized Shared Area
Randomized
Peeling
0.0
1.0
2.0
Selections
Target & Sister
Features
Confirm
Distance
Relative Angle
Neighbor Counts
Volume
Surface Area
Shared Area
Layer
Generation
Models
Show Similar Cells
SVM
RANDOM FOREST
KNN
NEURAL NETWORK
BAYESIAN
Accuracy / Recall(%)
100
80
60
40
20
0
94.29
61.34
94.2
48.77
93.28
50.23
92.41
63.59
92.05
83.89
SVM
Random Forest
KNN
Neural Network
Bayesian
Models Weight
Reset
Recalculate
SVM
RF
KNN
NN
Baye
0.00
0.10
0.20
0.30
0.40
0.50
0.60
N
N-1
N-2
N-3
...
1
Cell N-1
Cell N-3
Cell N-7
Cell N-8
Cell N-9
24
17
15
5
4
3
2
31
29
28
23
19
14
9
8
65
64
47
46
42
41
Interactions
Confirm
Wrong Children
New Sister
Supporting Cell
Predict New Level
Re-predict This Level

Evaluation Study

- Biologists appreciated the prediction results and their visualization.
- They thought LineageD+ could help save time and change the traditional approach they used in the assignment process.
- One biologist expressed that interacting with ML made her feel like she was discussing with the computer in making decisions.

Submission

Jiayi Hong, Ross Maciejewski, Alain Trubuil, and Tobias Isenberg. “Comparative Visualization of Machine Learning Predictions to Improve Human-AI Teaming on the Example of Cell Lineage”.

To TVCG, under review.

Comparative Visualization of Machine Learning Predictions to Improve Human-AI Teaming on the Example of Cell Lineage

Jiayi Hong, Ross Maciejewski, Alain Trubuil, and Tobias Isenberg

Abstract—We visualize the predictions of multiple machine learning models to help biologists as they interactively make decisions about *cell lineage*—the development of a (plant) embryo from a single *ovum cell*. Based on a confocal microscopy dataset, traditionally biologists manually constructed the cell lineage, starting from this observation and reasoning backward in time to establish their inheritance. To speed up this tedious process, we make use of machine learning (ML) models trained on a database of manually established cell lineages to assist the biologist in cell assignment. Most biologists, however, are not familiar with ML, nor is it clear to them which model best predicts the embryo's development. We thus have developed a visualization system that is designed to support biologists in exploring and comparing ML models, checking the model predictions, detecting possible ML model mistakes, and deciding on the most likely embryo development. To evaluate our proposed system, we deployed our interface with six biologists in an observational study. Our results show that the visual representations of machine learning are easily understandable, and our LineageD+ can effectively improve the assigning efficiency, reducing the time it takes to assign cell lineage and improving biologists' confidence in the ML models.

Index Terms—Visualization, machine learning, human-AI interaction.

1 Introduction

In biology, a plant cell (the *parent*) normally divides into two daughter (or *sister*) cells over time, and an embryo grows to eventually comprise hundreds of cells. To explore the history of an embryo's development, biologists utilize a 3D microscopy snapshot and assign sister relationships for every cell in the embryo. This is done backward across a series of snapshots, where biologists iteratively reason backward in time to arrive at the previous cell division stage. The datasets used in this process are extremely imbalanced because one cell can only have one correct sister cell, yet the cell usually has a dozen or more neighbors. As such, the manual assignment of the cell lineage for embryos of realistic sizes (several hundreds of cells) is extremely time-consuming

highly optimized models with high accuracy still have the potential to provide wrong predictions. In our cell lineage scenario, if a model wrongly predicts the assignments in the first few generations, the predictions for the following generations are almost certainly incorrect as well. Thus, biologists cannot exclusively rely on a completely automatic ML process. Instead, a human-AI teaming approach is preferred where experts can observe, control and update the labeling process. However, little work has concentrated on enhancing this *human-AI Teaming*[1] to assist experts in the decision-making process rather than focus their efforts on improving a given model's performance. To fill this gap, we visually represent the different ML predictions to assist the biologists, allowing them

05

Discussion

Main takeaways

Takeaways

01

Scientific Problem

Explored techniques for solving scientific problems.

02

Human-AI Teaming

Investigated the roles between humans and AI in task completion.

03

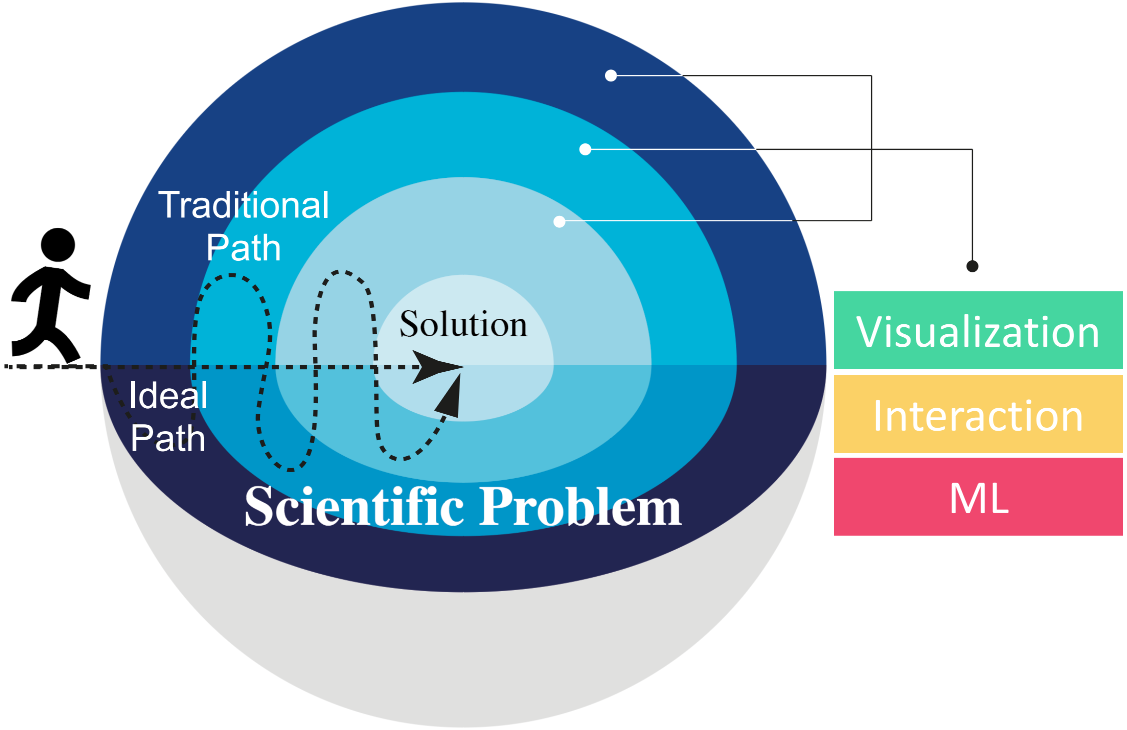
Link 3D and 2D

Explored ways to link two representations.

04

Application

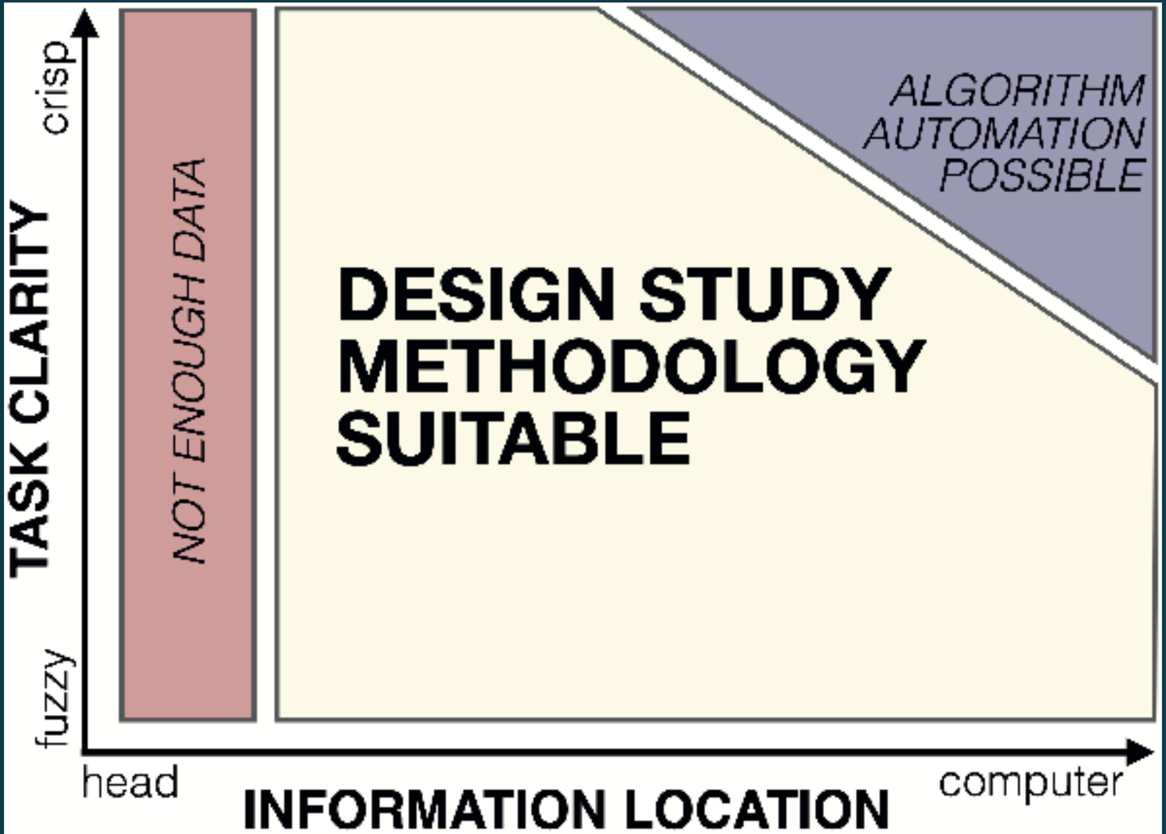
Developed and published our software for biologists to use.



Solving Scientific Problems

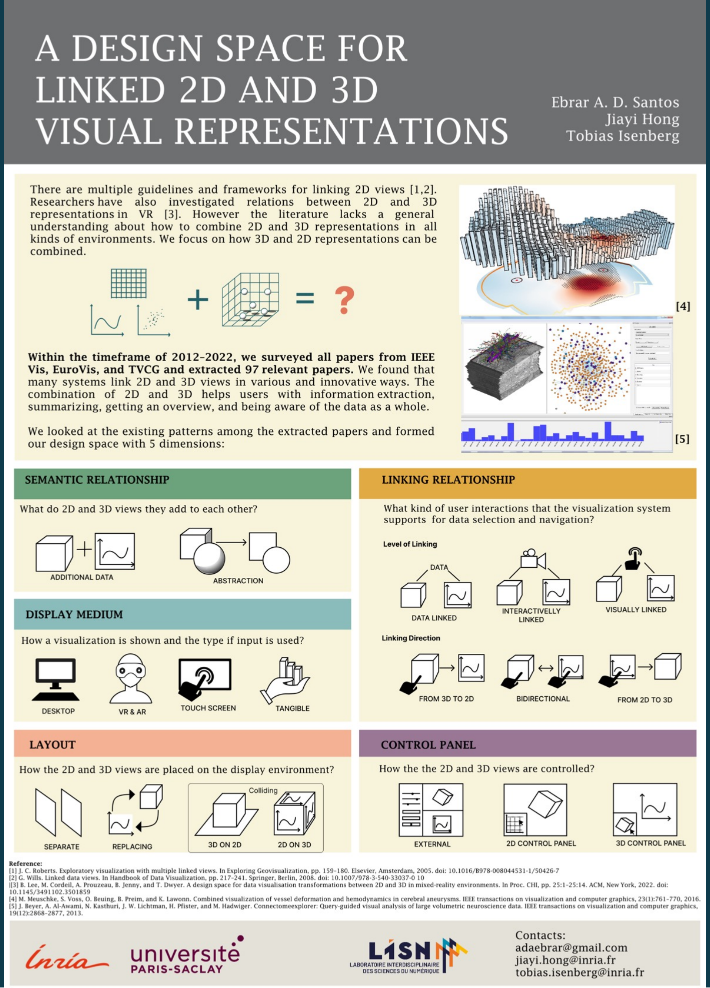
- Spherical Hairball

Human-AI Teaming



Inspired by [Sedlmair et al. 2012]

In the case of the boundary between two regions, we think there is no clear cut-off but, instead, a fuzzy transition zone where Human-AI Teaming is suitable.



Design Space

for Linked 2D and 3D Representations

[Santos et al. 2022]

- Semantic Relationship
- Display Medium
- Layout
- Linking Relationship
- Control Panel

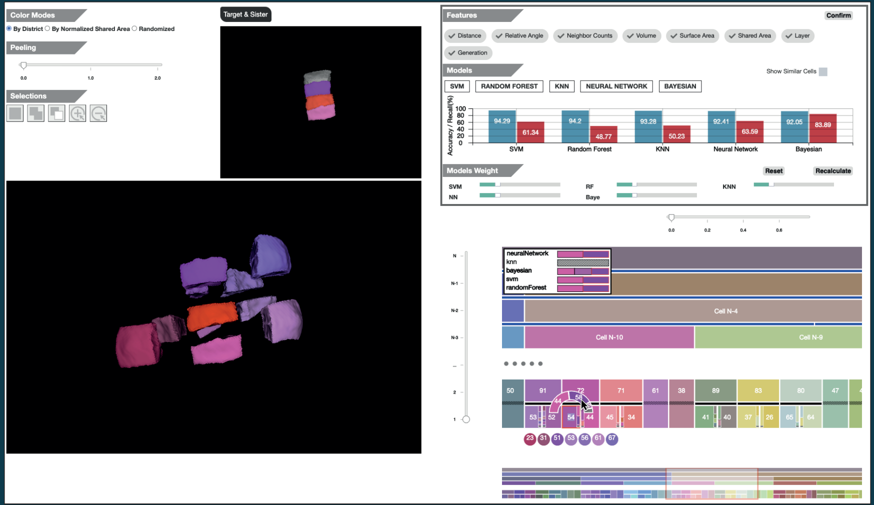
Application Contribution

We developed a tool for biologists to use when doing cell lineage.

LineageD

https://plantembryo.saclay.inria.fr/

LineageD+



https://plantembryoml.saclay.inria.fr/

Thanks!

https://jiayihong.info/

Conclusion

- Exploring scientific problems
- Human-AI Teaming
- Linking 3D and 2D representations
- Providing applications for biologists to do cell lineage