

CTransformer: Deep-transformer-based 3D cell membrane tracking with subcellular-resolved molecular quantification

Rigaudiere Z.L. Li

zelinli6-c@my.cityu.edu.hk

City University of Hong Kong <https://orcid.org/0000-0002-0985-8420>

Guoye Guan

Harvard Medical School and Dana-Farber Cancer Institute <https://orcid.org/0000-0003-4479-4722>

Xian Xiu

City University of Hong Kong

Dongying Xie

Hong Kong Baptist University

Yiming Ma

Hong Kong Baptist University <https://orcid.org/0000-0002-3646-3619>

Chenwei Wang

Centre for Intelligent Multidimensional Data Analysis

Sicheng You

Centre for Intelligent Multidimensional Data Analysis

Zhen Zhu

Centre for Intelligent Multidimensional Data Analysis

Darrick Lee

University of Edinburgh

Zirui Zhang

City University of Hong Kong

Zhuoheng Ran

City University of Hong Kong

Chao Tang

Academy for Advanced Interdisciplinary Studies, Peking University <https://orcid.org/0000-0003-1474-3705>

Jianfeng Cao

Harbin Institute of Technology (Shenzhen)

Zhaoke Huang

City University of Hong Kong <https://orcid.org/0000-0003-1837-9760>

Zhongying Zhao

Hong Kong Baptist University <https://orcid.org/0000-0003-2743-9008>

Hong Yan

City University of Hong Kong <https://orcid.org/0000-0001-9661-3095>

Article

Keywords: Cell Morphology, Molecular Quantification, Subcellular Resolution, E-cadherin, Cell Adhesion, Transformer, Generative Adversarial Network (GAN), Human-in-the-loop

Posted Date: March 4th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8944271/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

CTransformer: Deep-transformer-based 3D cell membrane tracking with subcellular-resolved molecular quantification

Abstract

Deep learning segmentation and fluorescence imaging techniques allow the cellular morphology of living embryos to be constructed spatiotemporally. These development processes involve numerous molecules distributed at the subcellular scale, such as cell adhesion (E-cadherin), which accumulate at cell-cell interfaces to regulate intercellular connections. However, quantifying molecular distributions within specific subcellular regions across the entire embryo, where cell movement and molecular redistribution occur rapidly, is challenging due to the need for simultaneous cell morphology reconstruction and lineage tracing due to photobleaching and phototoxicity. We report a transformer-based pipeline, *CTransformer*, that establishes a 4D cellular morphology map before the 550-cell (late) stage. *CTransformer* constructed 4D cellular morphology atlases, reaching 80% accuracy at the 550-cell stage. Through this advanced architecture, we use only one channel to reconstruct cell morphology and achieve cell tracing. With each cell's morphology as a reference, the distribution of specific molecules throughout the cell body and at cell interfaces can be quantitatively measured in another fluorescence channel. We apply this methodology to track E-cadherin during embryonic development of the worm *Caenorhabditis elegans*, from fertilization to gastrulation. Our results reveal that E-cadherin is tightly regulated across individual embryos, both within single cells and at cell-cell interfaces, displaying an anterior-posterior gradient and cell- and lineage-specific patterns. Furthermore, its spatiotemporal heterogeneity influences cell mechanics and embryonic morphogenesis, helping explain how *C. elegans* achieves stereotypical developmental patterns at cellular resolution. In summary, *CTransformer* offers a high-throughput approach for measuring molecular distributions at subcellular resolution, providing valuable knowledge for cellular and developmental biology, biophysics, and bioinformatics.

Keywords: Cell Morphology, Molecular Quantification, Subcellular Resolution, E-cadherin, Cell Adhesion, Transformer, Generative Adversarial Network (GAN), Human-in-the-loop

MAIN TEXT

Introduction

The spatial dynamics of molecules impact a lot of physiological functions from the subcellular to the organismic scale. For instance, in the mechanical sense, the life-dependent periodic exchange of actin filament (F-actin) between the cytosol and cortex of a cell determines its stiffness around cell division [1]. The epithelial cadherin (E-cadherin) located on the cell membrane affects not only the contact between neighboring cells but also drives collective cell sorting. In the chemical sense, the phase separation and movement of specific molecules inside a cell, such as the P granule that maintains the stemness in germline cells, is critical for specifying the stemness in its daughter [2]. Such polarized behavior is guided by the upstream reaction-diffusion molecule network that sets up gradients and cell polarization [3]. Moreover, the many important subcellular structures are in drastic spatial transformation that correlates to the physiological state of a cell, such as the centrosomes that drive cytokinesis. This numerous information at the subcellular scale is hard to monitor especially in a crowded multicellular system like a rapidly developing embryo, where the cell membrane morphology is usually hard to recognize and track simultaneously.

Cell morphology (shape, surface, and contacts) cannot be obtained simultaneously with other markers because of the number of channels needed and the problems of photobleaching and phototoxicity [4]. So specific expression of some properties, *e.g.*, membrane adhesion, can only be analyzed by approximate locations of nuclei rather than by precise cell morphology. If cell shapes and lineages can be traced with only one protein marker in *in vivo* embryos, photobleaching and phototoxicity will be lowered, keeping cells healthier and closer to “wild-type” behaviors. An extra channel can be used to observe other proteins, such as globular multi-functional protein family members, *e.g.* Actin, or cell adhesion proteins, *e.g.* *HMR-1*, a kind of E-cadherin, to coordinate developmental processes with the cell morphology map. Thus, conducting cell shape segmentation and cell identity tracing simultaneously via the 3D cell membrane tracking (for cell morphology map) is desired to quantify molecules at the subcellular scale. Although with tuned imaging balances, the advanced optical and deep learning segmentation algorithms are established [5-8], recognizing cell membrane signals and reconstructing morphology maps are far from perfect. Under normal circumstances, the maps are derived from membrane and nucleus fluorescently labeled images, giving the cell positions and detailed cellular shape. However, the two missions are usually realized by two complementary fluorescence channels: while the cell nuclei with condensed fluorescence intensity serve as the seed for cell identity tracing, they also guide cell membrane segmentation and assign cell identity to each cell region [6]. This way of occupying two channels limits the real-time monitoring of molecules of interest in the additional fluorescence channel. Therefore, simultaneously obtaining cell identities and membrane morphology with one channel is a technical solution not only making full use of image information but also providing one more fluorescence channel for monitoring a molecule. Recent works were performed on ascidian embryogenesis but still rely on manual correction and work on early developmental stages [9, 10].

In downstream biological analyses, incorrect cell shapes could trigger a chain of errors, which requires accurate cell morphology segmentation [4, 7, 9, 11-13]. With current adaptive scanning fluorescence microscopes (advanced hardware modalities) [14, 15], efforts to improve cellular recognition and segmentation focus on enhancing image clarity (super-resolution), resolving ambiguous signals, and generating instances (cell membrane recognition and instance segmentation). But it is not enough. The trade-offs between phototoxicity and image quality lead to laser attenuation, affecting imaging in ambiguous regions and introducing anisotropy, particularly along the axial direction, which impacts recognition and segmentation over time. Additionally, late-stage embryos contain numerous cells with irregular shapes and anisotropic sizes, making it challenging to accurately map single-cell morphology in dense cell structures. Even minor errors in cell membrane recognition can result in large areas of incorrectly connected or missing cells, and such errors may be exponentially amplified. While the cell shape segmentation using only cell membrane information is challenging, following cell identity tracing on top of it further requests its high accuracy as well as a robust tracing algorithm based on 3D cell regions, instead of the usual one on top of cell nuclei.

To keep embryos healthy, get cell identities, build membrane morphology, and discover molecular patterns, an accurate cell morphology map is indispensable; a nuclei-marker-independent method with short manual correction time is also missing, especially at the late-cell stage (over 550 cells). By building a cell morphology map with nuclei-independent images, it should be possible to build an expression map along a single-cell shape. In previous studies, other gene-markers or proteins were estimated along the cell nuclei map only [16, 17]. Cell morphology (shape, surface, and contact) cannot be derived simultaneously with other markers because of the number of channels, photobleaching, and phototoxicity [4, 7]. So, previously specific expressions are analyzed only along with approximate nuclei locations rather than on precise cell morphology [16, 18]. By generating both the cell shape and the lineage tracing with only one protein for *in vivo* embryos (no nuclei fluorescence staining), first we can lower the photobleaching and phototoxicity effects to keep healthier and more “wild-type” embryos; second, we can use the extra channel to observe other proteins based on accurate cell shape, like globular multi-functional protein family, e.g. Actin, or cell adhesion protein, e.g. *HMR-1*, a kind of *E-cadherin* (Figure 1), to observe some developmental processes with cell morphology map. We want that actual and real expressions on the segmented shape, cell surface, and cell-cell interaction contact in *in vivo* embryos can be quantitatively calculated and analyzed, allowing researchers to carry out previously unachievable experiments to discover regulations between cell shape and other expressions. Like cell adhesion [19, 20], a fundamental property in multicellular development, whose surface-wise and contact-wise patterns in the living structural systems can/are not be studied.

In this paper, we introduce a generalizable and adaptive integrated DNN pipeline, *CTransformer*, for 4D image cellular segmentation (*TUNETr* Module), nuclei-independent lineage tracing (*m2nGAN* Module), and shape-expression analyses (*MolQuantifier* Module). With *CTransformer*, the actual expression of specific

genes and proteins associated with segmented shape, cell surfaces, and cell-cell interaction contacts in *in vivo* embryos can be quantitatively calculated and analyzed. This allows previously unachievable experiments to discover regulatory patterns between cell shape and other cell membrane features like cell adhesion [19, 20], a fundamental property in the development of multicellular organisms, where surface shape and contact patterns are informative. One channel is used to segment every single cell’s shape through fluorescent membrane signals; centroids of the segmented shapes are used to generate images of pseudo-nuclei for efficient lineage tracing. Another channel measures the intensity of each cell’s adhesions, allowing the discovery of patterns (Figure 1). The DNN model of *CTransformer* for segmentation by morphology is a semi-supervised Topology-constraint U-shape Nucleus-prompting Euclidean distance transform Transformer (*TUNETr* Module). To ensure that *TUNETr* performs well with limited GT training, *TUNETr* integrates the relative positional attention mechanism of transformers [21], prior inductive bias knowledge, and image processing technologies in a U-shaped network, topology-constraint loss function [22, 23]. Boundary-aware semi-supervised learning and synthetic GT data are utilized to solve the challenges in failure to segment axial and unseen images. These modules together form a human-in-the-loop pipeline to automate cell segmentation based on morphology. The potential of *CTransformer* for 4D *in vivo* cellular segmentation with *C. elegans* (Movies S1 and S2), showing time-lapse morphology map. Moreover, A membrane to nucleus Generative Adversarial Network (*m2nGAN* Module), a generative model to artificially generate images of pseudo nuclei from cell shapes is developed to allow tracing of cell lineages from only membrane images. *CTransformer* was applied to the segmentation of *C. elegans* live embryos, previously imaged and annotated by 5 biological experts, under 3 different imaging conditions, with 6279 and 2339 cell objects, for training and evaluation, respectively. Segmentation of the *in vivo* embryos was evaluated by multiple metrics, in comparison with 9 existing DNN-based methods, on 30509 cell regions, which demonstrated its superior performance. The images of pseudo nuclei for *C. elegans* were tested and showed effectiveness in tracing cell lineages up to 550-cell stage embryos. A user-friendly and feasible scripts are provided for users, allowing researchers to run segmentation and visualization of the morphology map in ITK-SNAP-CVE [4, 24]. The proposed pipeline could pave the way for tracing cell lineages using membrane images, and enable discoveries of more transcriptomic, e.g. adhesion expression, shape-related molecule patterns.

Results

***CTransformer* pipeline is designed to track 3D cell membrane in fluorescence images, enabling both accurate cell morphology segmentation and cell lineage tracing.**

CTransformer is a fully automated and integrative experimental-computational framework designed to track time-lapse 3D cell membranes and reconstruct both cell morphologies and lineages across entire samples, using only the experimental fluorescence channel of cell membranes, which quantifies spatiotemporal dynamics of molecules of interest on top of the reconstructed cell morphologies and lineages simultaneously using another fluorescence channel (Figure 1). In the first channel, 3D + t images of cell membrane fluorescence are subjected to morphology segmentation (*TUNETr* Module - a deep neural network or DNN for voxel-wise segmentation of cell membranes) (Figure 1a; Figure S2); then the segmented morphologies are processed to generate cell nuclei pseudo-fluorescence for lineage tracing, without reliance on the experimental fluorescence channel of cell nuclei as before. We use the proposed *m2nGAN* Module - a generative adversarial network (GAN) transforming membrane-segmented images into nuclei-mimicked images (Figure 1b) [25-28]. Building upon the reconstructed whole-sample cell morphologies and lineages, the additional channel(s) of 3D + t images of labeled molecule(s) provide its quantitative distributions of molecular signals at any spatial and time points (*MolQuantifier* Module - an adjustable combination of molecular measurements specialized for spatial regions, including but not limited to those accumulated on specific cell surfaces and cell-cell contact interfaces). Altogether, the full utilization of multiple fluorescence channels enables *CTransformer* to automatically delineate the complicated spatiotemporal dynamics of entire biological samples, spanning from molecular, subcellular, cellular, to multicellular scales.

To satisfy the segmentation accuracy requirement for *CTransformer* construction, we first established a well-organized dataset for multiple *in vivo* *C. elegans* embryos documented with a unified data format. This dataset contains the 3D + t images (a total of 16,922 dual-channel volumetric images) of 34 compressed embryos (*incl.*, 29 wild-type samples developing normally and 5 RNAi-treated samples developing abnormally) and 14 uncompressed (*incl.*, 9 wild-type samples developing normally and 5 RNAi-treated samples developing abnormally) embryos for establishing the *TUNETr* Module and *m2nGAN* Module. Those embryo samples are fluorescently labeled by green fluorescent protein or GFP to label cell nuclei (stimulated by laser beam at 488 nm) and mCherry to label cell membranes (stimulated by laser beam at 594 nm), covering *C. elegans* embryogenesis from 2- to >550-cell stage, namely, from the completion of the first cell division to the last round of most cell divisions. While a part of the embryo samples are adopted from previous public datasets for enlarging training data pool and are subject to new segmentation/tracing with higher quality or yet to be done [4, 7], another part are newly published in this paper to test *CTransformer*'s general applicability (Table S1). Last but not least, a new strain was newly crossed with mCherry labeling cell membranes and GFP labeling

adhesive protein HMR-1/E-cadherin, a total of 20400 dual-channel images, producing three additional embryos that were imaged from fertilization to gastrulation, for establishing the *MolQuantifier* Module.

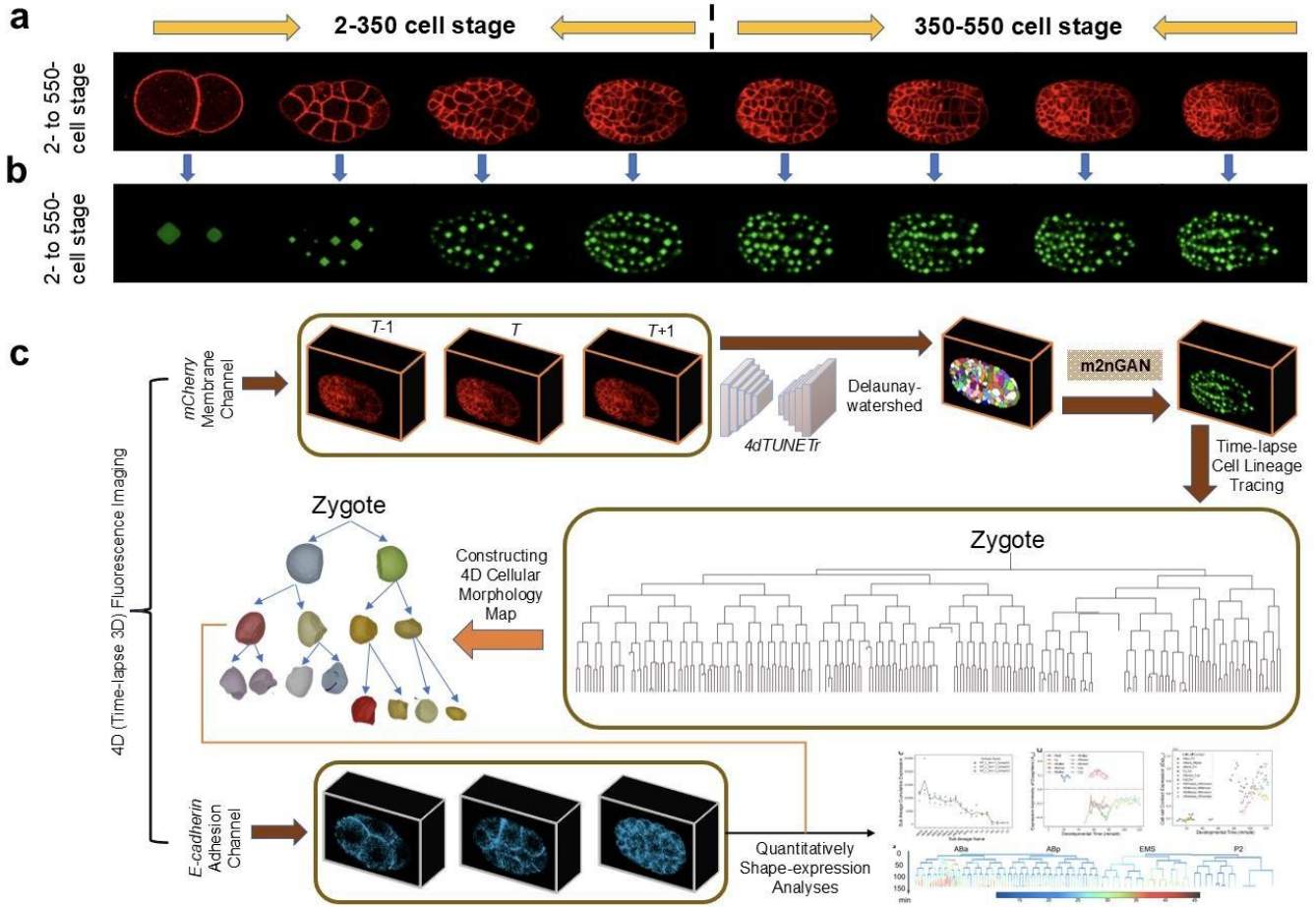


Figure 1 | Experimental-computational design of the integrative *CTransformer* framework, exemplified by its application to worm *C. elegans* embryogenesis. (a) *in vivo* collection of 3D + *t* images about fluorescence labeling cell membranes (displayed with 2- to 350-cell stages on the left and 350- to 550-cell stages on the right, when the fluorescence signal becomes progressively ambiguous and difficult to recognize). (b) *in silico* generation of 3D + *t* images about pseudo-fluorescence mimicking cell nuclei corresponding to the fluorescence labeling cell membranes in (a). (c) *CTransformer* workflow from cell membrane fluorescence images to cell nuclei pseudo-fluorescence images, which are subsequently used for cell morphology segmentation (*TUNETr* Module) and cell lineage tracing (*m2nGAN* Module), respectively, ultimately enabling whole-sample, subcellular-resolution quantification of molecules of interest (exemplified by the adhesive protein HMR-1/E-Cadherin) across space and time (*MolQuantifier* Module).

***CTransformer* segments cell morphologies accurately, consistently performing well even under conditions of large cell number and high cell density.**

The first module of *CTransformer* is *TUNETr*: a scalable and high-capacity vision model, inspired by SwinUNETR [29, 30], specifically tailored for voxel-wise segmentation in large-scale embryonic imaging datasets. It is designed to capture complex 3D morphological features while maintaining robustness across a wide range of developmental stages and imaging conditions. The model is initialized with *iTUNETr*, a backbone pre-trained on a curated set of manually annotated 3D volumes and progressively refined through two specialized variants: *sTUNETr* (for static 3D snapshots) and *4dTUNETr* (for continuous time-lapse data). These variants are trained using a semi-supervised learning framework that incorporates edge-aware supervision, pseudo-label propagation, and topology-informed structural priors (Figures S1 and S2a). This human-in-the-loop and multi-stage training strategy (Figure S3), anchored by expert-curated cell lineage trees, enables precise boundary delineation and ensures robust generalization across diverse spatiotemporal imaging conditions (Figure S4).

The potential of *CTransformer* for accurate cell segmentation was tested on the dataset (named “*3D Shape Evaluation*” dataset accessible via <https://doi.org/10.6084/m9.figshare.27085657>), which includes manually annotated ground truth (GT) from 16 distinct 3D volumetric images with 2,339 3D cell objects [4]. These annotations were curated to represent significant developmental milestones, providing a robust basis for evaluating segmentation capabilities prior to the 550-cell stage. *CTransformer* was compared with a series of state-of-the-art algorithms, *3DUNet++* [31], *VNet* [32], *SwinUNETR* [29], *Cellpose3D* [33], *StarDist3D* [34], and *Cshaper++* (*CShaper*) [12]. Each method was tested under identical conditions, ensuring a fair and unbiased comparison; their performance was evaluated by the Dice Score, Hausdorff Distance, and Jaccard Index, and compared with each other.

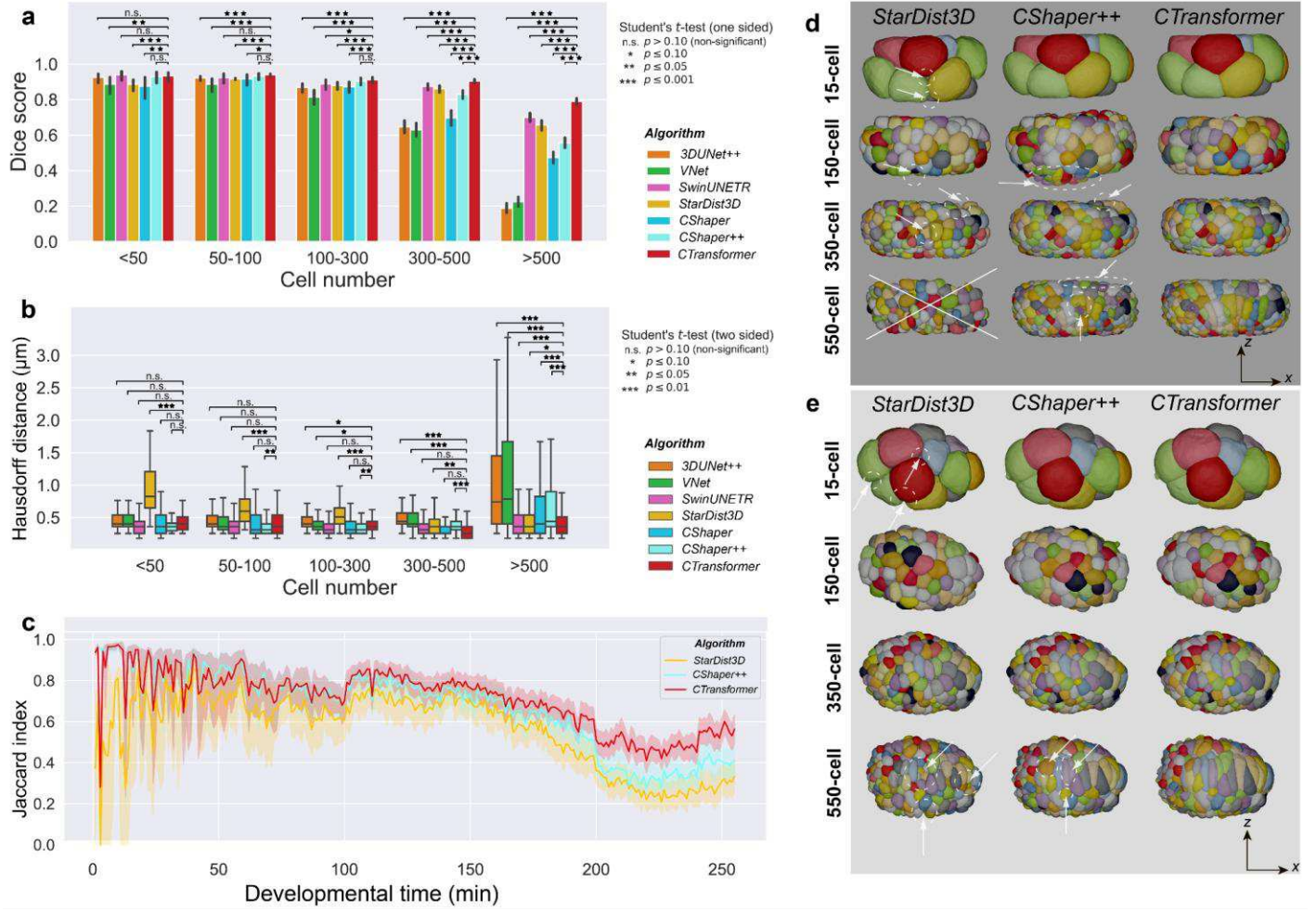


Figure 2 | Outperformance of *CTransformer* compared with other state-of-the-art algorithms, in terms of segmentation quality. (a) Dice score over developmental stage defined by cell numbers, demonstrating *CTransformer*'s outperformance at late developmental stages. **(b)** Hausdorff distance at different developmental stages defined by cell numbers, demonstrating *CTransformer*'s outperformance at late developmental stages. **(c)** Jaccard index over developmental stages defined by time point, demonstrating *CTransformer*'s outperformance at late developmental stages. **(d)** Segmentation results of a compressed embryo at its 15-, 150-, 350-, and 550-cell stages (illustrated with the embryo sample WT_C_Sample2). **(e)** Segmentation results of an uncompressed embryo at its 15-, 150-, 350-, and 550-cell stages (illustrated with the embryo sample WT_Sample1).

Dice Score is a statistical and pixel (voxel) based measure that assesses the similarity between predicted cells ($Cell_{pred}$) and ground truth cells ($Cell_{gt}$). It is defined as twice the area of overlap between the two segmentations divided by the total number of voxels in $Cell_{pred}$ and $Cell_{gt}$:

$$Dice = \frac{2 \times |Cell_{pred} \cap Cell_{gt}|}{|Cell_{pred}| + |Cell_{gt}|} \quad (1)$$

CTransformer achieved an overall value of 0.908 ± 0.003 , the greatest accuracy in embryonic cell segmentation and 0.02 higher than the second-best method, *SwinUNETR*, which achieved a Dice score of 0.888 ± 0.006 (Figure 2a). This metric, indicative of the overlap between predicted and actual cell segmentations, highlights

CTransformer's effectiveness in capturing cell morphology (Table S2 and S3). While *CTransformer* has comparable performance with *3DUNet++*, *CShaper++*, *SwinUNETR*, it significantly outperforms all other algorithms from 300- to beyond 550-cell stage.

Hausdorff Distance measures the greatest distance from all voxels in the surface of $Cell_{pred}$ to their closest voxel in the surface of $Cell_{gt}$. It is calculated as:

$$H(Cell_{pred}, Cell_{gt}) = \max(h(Cell_{pred}, Cell_{gt}), h(Cell_{gt}, Cell_{pred})) \quad (2)$$

where $h(Cell_{pred}, Cell_{gt}) = |a - b|$ and $h(Cell_{gt}, Cell_{pred}) = |b - a|$. *CTransformer* had a score of $2.591 \pm 0.187 \mu\text{m}$ (Table S2), indicating minimal boundary discrepancies and higher fidelity in representing cell shapes. This performance was generally better than other methods, with *SwinUNETR*, the second best, scoring $2.598 \pm 0.339 \mu\text{m}$ (Figure 2b).

The **Jaccard Index** (JI), also known as Intersection over Union (IoU), was computed for successfully segmented cells and averaged across all cells. It is demonstrated as:

$$\text{Jaccard Index} = \frac{\text{True Positive (TP)}}{\text{TP} + \text{False Positive (FP)} + \text{False Negative (FN)}}$$

These metrics were selected for their ability to thoroughly evaluate the performance of segmentation across temporal sequences and varying cell stages (Figure 2c; Table S4), reflecting the complexity of embryonic development. All state-of-the-art methods experienced a decline in accuracy over time (~359 minutes from the 4-cell stage to the 589-cell stage). However, *CTransformer's* degradation was the smallest, maintaining the highest accuracy across different developmental stages. It achieved a Jaccard Index of 0.715 ± 0.02 (Table S5), outperforming *StarDist3D* (0.563 ± 0.033 ; Table S6) and *CShaper+* (0.673 ± 0.037 ; Table S7).

The time-lapse curves of two embryos further demonstrated *CTransformer's* temporal robustness and consistency throughout embryonic development (Figure 2, d and 2e, and S5).

All the above three metrics coherently highlight *CTransformer's* unique segmentation performance, especially for **conditions of large cell number and high cell density** (Figure S5).

Overall, *CTransformer* outperformed other deep learning-based methods by over 10% for construction of cell morphology maps, as shown at five key developmental stages. These results were consistent across various imaging conditions (two confocal systems), culture situations (compressed or uncompressed), and different scanning depths (68, 70, and 92 layers).

In addition to single time-point evaluations, time-lapse assessments of embryonic development were assessed to evaluate overcoming challenges associated with constructing comprehensive $2D + t$ ground truths. The **Time-lapse 2D Evaluation** dataset (<https://doi.org/10.6084/m9.figshare.2708565>) consisted of two embryos with a total of 460 images and 30,509 cell regions. Annotating the central 2D YZ plane cut from over 200 volumetric series gave a robust way to evaluate segmentation performance over time. *CTransformer*

demonstrated strong temporal robustness, as evidenced by its superior Jaccard Index compared to established methods like StarDist3D and *CShaper++*.

Qualitative comparisons between StarDist3D, *CShaper++*, and *CTransformer* are presented (Figure 2, d and e, S5 and S6). Regardless of compression conditions, cell stages (350- or 550-cell), or imaging planes (axial or sagittal), *CTransformer* consistently made fewer visual errors.

With the above results, in summary, the **TUNETr** Module is an effective tool for segmentation of live-cell embryos by 4D morphological mapping. Its proficiency across various metrics and conditions represents a significant advance for the accurate portrayal and understanding of cellular dynamics. *CTransformer*'s application in biological studies will open avenues for future exploration into its potential enhancement and broader implications for 3D + t biological imaging.

With only membrane FC images, *CTransformer* tracks both cell morphology and lineage across multiple dimensions throughout embryonic development with high fidelity.

In this section, achieving low rates of cell loss during the fluorescent imaging and optical scanning of 3D embryos is described. 4D images may encounter intrinsic light heterogeneity and weak signals, as in the inner and middle cell layers due to larger inner cells and less dense expression of signaling proteins (*mCherry*) (Figure 1a). This anisotropy can lead to failures in segmentation and cell loss. Relative to other methods (Figures 3, a and b, and S6), *CTransformer* showed lower anisotropy in brightness, more balanced signal strength across different cell layers, and better recognition of 3D membrane at attenuated laser intensities. *CTransformer* effectively resolved weak boundaries (at the exterior of the embryo) by focusing on thickening outer membranes in the training strategy (stronger signals in boundary at Figure S7). The topological loss helped connect incomplete and invisible membranes by learning the embryo's membrane skeleton in a topological latent space. Here, for this intermediate step in recognition of membranes, the reconstruction of 3D cell membranes in ambiguous images was improved for consecutive single-cell segmentations, especially for late-stage embryos with small cells.

The performance of *CTransformer* surpassed that of other methods in reconstructing single-cell morphology with minimal cell loss across time, tissue type, and lineage (Figure 3, c to e). The evaluation was conducted on 1 wild-type compressed and 1 uncompressed *C. elegans* embryo, with their last developmental times being 240 and 280 minutes, respectively. The rate of cell loss was determined by comparison against manually traced nuclei numbers to assess the proportion of missing cells in predictions. *CTransformer* successfully generated 96.5% of cell objects, while StarDist3D and *CShaper++* constructed 62.0% and 83.7% of cells in the two evaluated embryos (Table S8).

CTransformer could reconstruct skin cells to the 550-cell stage without errors but there were some errors after that (Figure 3c) [35], with low loss rates (around 1%), achieving a tenfold improvement over other

methods. The average loss rate of cells of extremely small size (less than $10\mu\text{m}^3$), like apoptotic cells, was reduced to $\sim 20\%$ (Figure 3d). Different cells could be better constructed in the morphological map by *CTransformer*, enabling more feasible studies of biological processes, like apoptosis, with support from the more accurate cell-cell contact atlas. In lineage assignment, *CTransformer* dealt with deformed and small sizes with lower loss rates, less than 15% for progeny of the AB and C lineages (Figure 3e, Table S9).

Moreover, we evaluate the reproducibility of the cell surface area of the three *HMR-1* embryos from the segmented cells. The same cells in different embryos have similar cell surface area, which obeys the development (Table S10); the consistent cell shape segmentation and adhesion results show the accuracy of capacity of *CTransformer*'s morphology map construction.

Cells segmented by *CTransformer* exhibited better consistency in recognition of the cells and their shapes over time in the real embryonic cell map.

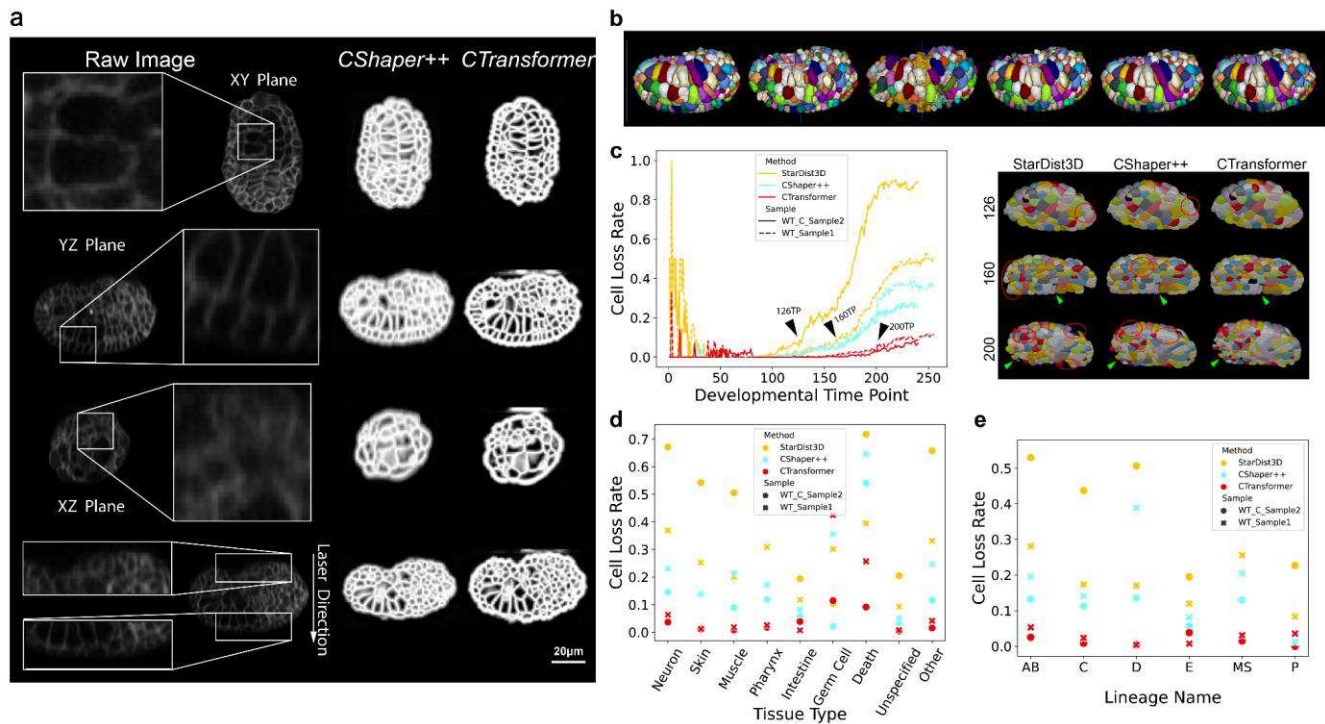


Figure 3 | Overall 3D cell membrane tracking performance in terms of image enhancement, cell morphology reconstruction, and cell lineage tracing. (a) Membrane recognition at ~ 550 -cell stage. While cell blurring occurs in different regions due to different reasons (e.g., at the axial imaging plane, the first and fourth rows, ambiguous areas are prevalent because of isotropic resolution, light scattering, and low resolution), *CTransformer* enhances the image quality substantially and better than *CShaper++*. (b) Cell morphology reconstruction at different developmental stages. (c) Cell loss rate in compressed and uncompressed embryo samples, regarding developmental time. (d) Cell loss rate in compressed and uncompressed embryo samples, regarding cell fates. (e) Cell loss rate in compressed and uncompressed embryo samples, regarding cell lineages.

With the *m2nGAN* Module, *CTransformer* traces cell lineages accurately by generating pseudo cell nuclei fluorescence images near the real ones.

The second module of *CTransformer* is *m2nGAN*: a membrane-to-nucleus generative adversarial network that synthesizes high-fidelity pseudo-nucleus images from membrane-only 4D images, thereby enabling nucleus-free lineage tracing (Figure 1c). The architecture comprises dual generators (G_A , G_B) and dual discriminators (D_A , D_B) that are jointly optimized to reconstruct nucleus morphologies with high contrast and biological plausibility. Generator G_A is conditioned on the precise cell boundaries provided by *4dTUNETr* and outputs synthetic nuclei that surpass conventional (raw real) nucleus labels in signal clarity, reducing background noise and improving the visibility of subnucleus structures (Figure 1b and S8). This approach not only bypasses the need for nucleus labeling in live imaging but also enhances lineage reconstruction accuracy in densely packed cellular environments.

To quantitatively evaluate the pseudo generative (results of *m2nGAN*) and real raw nuclei images (second row in Figure 1c), 2 image-wise metrics, peak signal to noise ratio (PSNR) based on the mean square error (MSE) and signal-to-noise ratio (SNR) were used, as well as the standard criteria, precision, recall, and accuracy, on the “**Lineage Tracing Evaluation**” dataset (accessible via <https://doi.org/10.6084/m9.figshare.27085657>). This dataset includes 3 time series, 1 compressed and 2 uncompressed, of embryos with 625 embryonic volumes, as well as all their edited nuclei (Figure S8; Table S11); all of their maximum imaging time points are up to the 550-cell stage. The effectiveness of *m2nGAN* at constructing generative pseudo nuclei images to establish cell lineages is demonstrated (Figure 4).

PSNR is defined as

$$PSNR = 20 \times \log_{10} \left(\frac{MAX_f}{\sqrt{MSE}} \right), \quad (4)$$

where MAX_f is 255 and the MSE is formulated as $MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \left(I_{gt}(i, j), I(i, j) \right)^2$.

SNR is calculated as $SNR = \frac{P_{signal}}{P_{noise}}$, defined for images of nuclei measured in this study. P_{signal} and P_{noise} represent the total power of signals and noise, respectively, as measured by the accumulated intensity of the pixels inside and outside the nuclei. Signal refers to the voxels located within the cell nuclei regions identified in GT images and noise consists of voxels outside those regions.

Precision, recall and accuracy are formulated as:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}, \quad (5)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}, \quad (6)$$

$$Accuracy = \frac{True\ Positives}{True\ Positives + False\ Positives + False\ Negatives}. \quad (7)$$

At every time point, they are computed according to the cell identities within the time series (Figure S9), as has been done widely previously [36-38]. Error numbers are compared and categorized into 5 types (Figure 4b), i.e., division error and tracing error [39]. Qualitative assessments of the whole lineages and single imaging slides (intensity heterogeneity) and a comprehensive lineage-wise evaluation of the generative images, from the perspective of tracing of the cell lineage by biologists, are also provided.

Additionally, qualitative traced lineage tree comparisons between raw (real fluorescence) and generative images also demonstrated the effectiveness of *m2nGAN* (Figures 4c, S10 to S13), highlighting the following 3 advantages of *m2nGAN*,

- (1) In top, middle, and bottom focal planes, the pseudo-nucleus exhibited stronger and more uniform signals, with reduced background artifacts.
- (2) In early to late embryonic stages, pseudo-nucleus images maintained consistent contrast and structural clarity (Figure 4d), supporting stable lineage inference over time.
- (3) Lineage trees constructed from generative images were complete and more consistent with biological expectations.

For a more comprehensive evaluation, we conducted the quantitative image-wise and standard measurements to show that *m2nGAN* generated images of nuclei are better than real (raw) images they have higher accuracy and SNR, lower MSE and error numbers than real nuclei images (Figures 4a, S8, and S9) for cell lineaging. Construction and manual curation of cell lineages became easier and more efficient with fewer tracking and tracing errors qualitatively in generative images (Figures 4b and S10).

With above evaluated results, generated images (the results of *m2nGAN*), give equally good or even better signals of nuclei and image quality to allow lineage tracing to build cell morphological maps. This phenomenon is mainly caused by the early lineage tracing errors and the subsequent propagation of real nuclei images.

As the *m2nGAN Module* allows cell nucleus position remodeling and lineage based on only the cell membranes, dynamics of other molecules of interest, not only for the previously nuclei-relied ones (e.g., promoter and chromatin activities [40, 41]) but also for the ones displaying spatial dynamics beyond the nucleus (e.g., *HMR-1*/Ecadherin located on the cell membrane surface/interface), can be monitored with an additional fluorescence channel on top of the cell morphology and cell lineage.

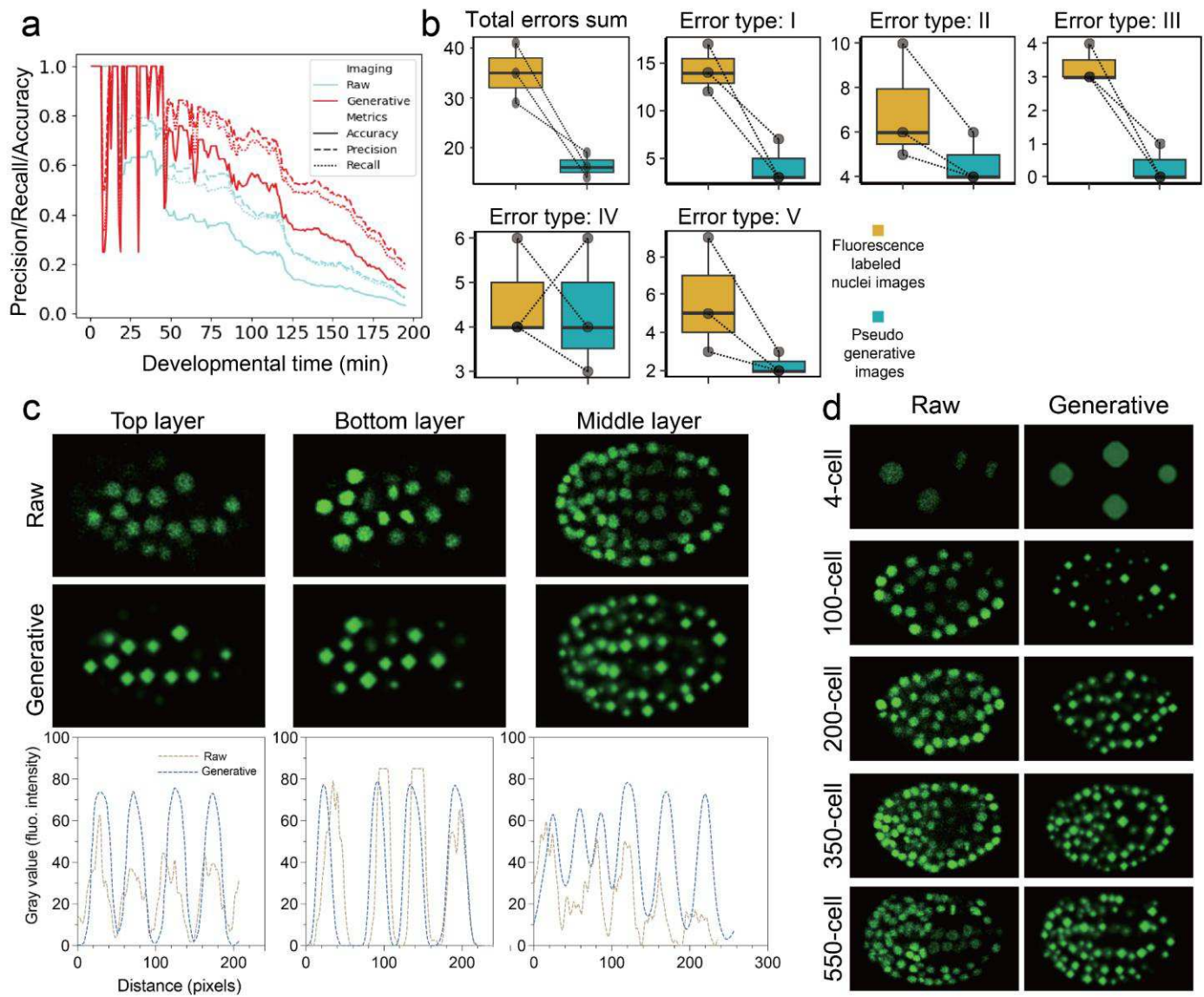


Figure 4 | Comparable real (based on cell nuclei fluorescence) and generative (based on cell membrane fluorescence and pre-trained *m2nGAN* module) images for cell lineage tracing. (a) The precision (dashed line), recall (solid line), and accuracy (dotted line) curves of cell lineage tracing with real (blue) and generative (red) images. (b) The number of errors produced by real (yellow) and generative (cyan) images, measured at the ~160-cell stage from three individual embryos samples. A dashed line connects the error counts for the statistical results in each embryo of raw (real) and generative (pseudo) images. (c) Smoother, more even and more distinct image signal distribution in generative images than real images, displayed with different focal planes. The intensities of fluorescence of raw (blue) and generative (red) images, measured in the whole image. (d) Smoother, more even, and more distinct image signal distribution in generative images than real images, displayed with different developmental stages.

With the *MolQuantifier* Module, *CTransformer* automatically quantifies spatiotemporal dynamics of a labeled molecule across subcellular to multicellular scales with the spare fluorescence channel.

The third module of *CTransformer* is *MolQuantifier*: a set of specialized algorithms designed for **shape-based representation analysis**, which plays a critical role in interpreting complex 4D cellular morphology (Figures 5 and 6; Table S12). These algorithms extract high-dimensional geometric features from segmented cell membranes, including shape descriptors, spatial orientation, and relative positioning within the embryo context. By analyzing how these shape features evolve over time and across lineages, the pipeline enables the identification of developmental patterns such as coordinated cell elongation, directional growth, and spatial organization of cell assemblies. Furthermore, by integrating adhesion information from the second imaging channel, the system can correlate morphological changes with cell–cell contact dynamics, offering insights into how physical interactions drive morphogenetic processes.

CTransformer's remarkable capability to track 3D cell membranes independently of cell nuclei offers a chance to quantify spatiotemporal dynamics of any molecule of interest through fluorescent labeling in a separate channel (Figures 2, 3, 4). This capability is technically important because the number of distinct fluorescent markers and non-overlapping imaging channels in microscopy is often limited [42, 43]. Since cell morphologies and cell lineages across the entire organism can be reconstructed solely from cell membrane fluorescence, such quantification can reach subcellular resolution. In other words, this methodology allows for monitoring how molecular distributions vary within and across 3D cell bodies over time accurately. To demonstrate this idea, we utilized the protein *HMR-1*/E-cadherin, which is well-known to accumulate at cell interfaces and mediates adhesion between neighboring cells (Figure 5a) [19, 44]; as an evolutionarily conserved protein from worms to humans, it has been implicated in multiple developmental processes, including body axis establishment, germ layer segregation, intestinal formation, and other key morphogenetic processes [45, 46]. Fluorescent intensity values are used as a proxy for the physical surface adhesion in the following quantitative analyses (Figures 5 and 6, a to d).

In practice, we crossed *C. elegans* strains with mCherry labeling cell membranes and GFP labeling *HMR-1*/E-cadherin protein, respectively, to generate a new strain ZZY0868 [47]. Using confocal microscopy equipped with two fluorescence channels, one channel records the cell membrane signal for cell morphology segmentation and cell lineage tracing, while the other channel records the cell membrane signal *HMR-1*/E-cadherin protein for spatiotemporal dynamic quantification. As cell membrane fluorescence successfully supports cell morphology segmentation and cell lineage tracing, the *HMR-1*/E-cadherin protein signal can be mapped onto the resulting cell morphology and lineage information, reproducibly across three embryo replicates developing up to the gastrulation stage (Figure 5a). In particular, fluorescence of a labeled molecule

(represented by *HMR-1*/E-cadherin protein here) is extracted for the entire cellular surface as well as for each intercellular interface that composes it, yielding average intensity values over a thin 3D layer of the surface or interface (Methods). Thus, fluorescence intensity quantified over cellular surface exhibits a lineage-resolved distribution of the labeled molecule, opening a new data dimension that complements and extends previous lineage-resolved studies of properties such as cell cycle length and division timing [48, 49], cell shape descriptors [50, 51], gene expression [52, 53], and chromatin accessibility (Figure 5b) [40, 54].

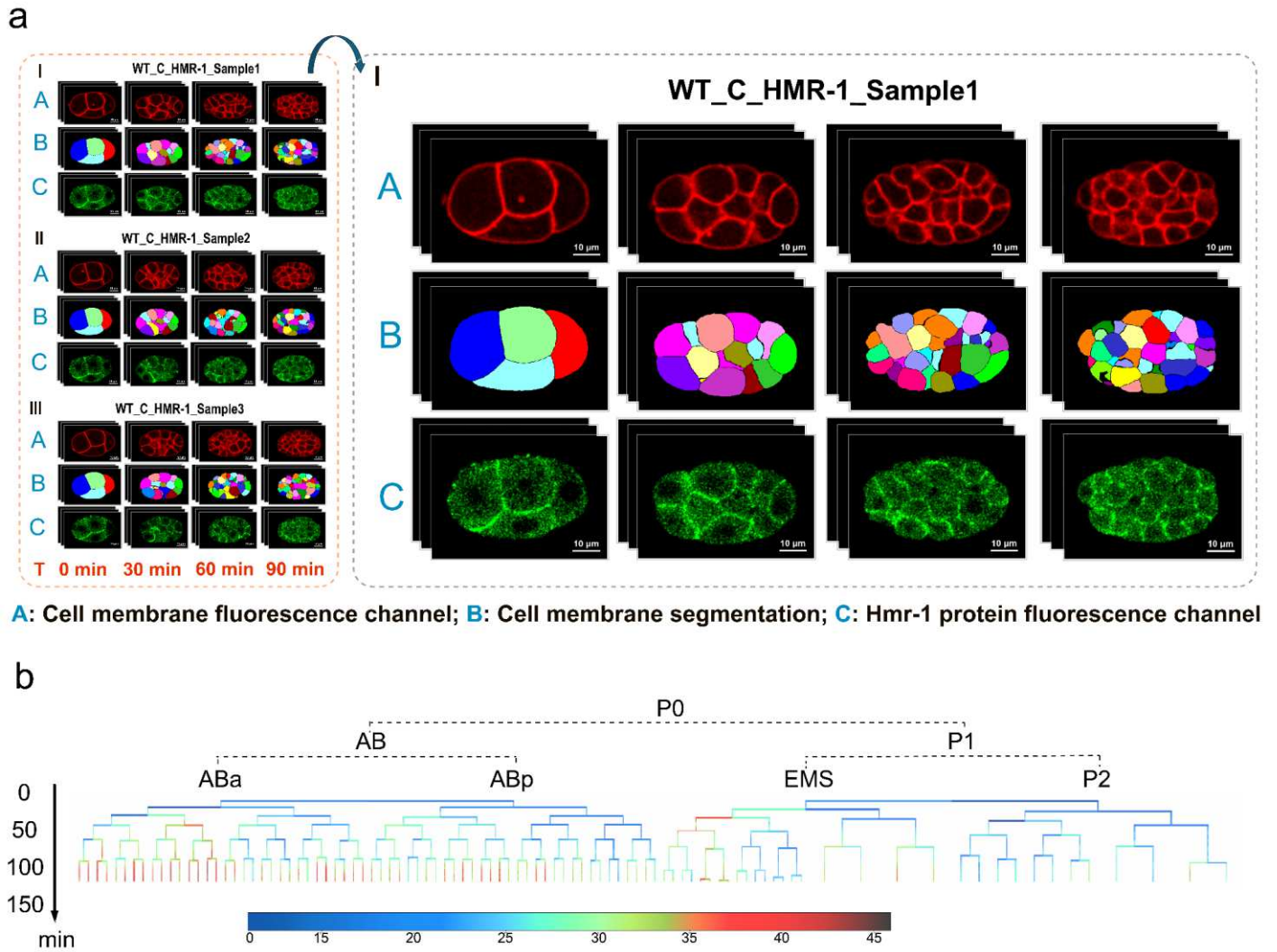


Figure 5. HMR-1/E-Cadherin regulated adhesion quantification demonstration. (a) The time-lapse cell adhesion quantification demonstration: cell membrane staining, cell morphology map establishing, temporal-spatial cell adhesion quantification. (b) Cell adhesion, expressed in units ($0.0324 \mu m^2$) of surface area, over a complete lineage tree. Lineage-wise cell surface *HMR-1*/E-Cadherin expression tree is demonstrated, while cell-wise and fate-wise adhesion distributions are presented.

HMR-1/E-Cadherin quantification reveals various regulated adhesion behaviors with specificity to space, time, lineage, cells, and cell-cell contacts.

First, we evaluate the distribution of the accuracy of cell adhesion measurement results (Figure 6a). The reproducibility of adhesion demonstrates the reliability of *MolQuantifier* Module (Table S10). In the lineage-scale quantification, the AB lineage exhibits stronger adhesion than P1, which may account for previous observations that AB cells are more regular than P1 cells (Table S12). While the AB lineage occupies the anterior and the P1 lineage the posterior, the particularly stronger adhesion (compared to ABp sublineage) observed in the more anterior ABa sublineage within AB lineage further; This anterior preference, observed across all 18 sublineages (Figure 6b), further suggests that the spatial heterogeneity in adhesion is biased along the anterior-posterior axis as a global manner. Whether this results from internal lineage differentiation or external diffusive signaling (*e.g.*, Wnt signaling) warrants further experimental investigation [4, 9]. In consideration of the stereotypical bias of cell fate in AB and P1 lineages, whether these higher surface adhesion areas contribute to the further differentiation of pharyngeal and neuronal tissues concentrated in AB lineage, or in contrary, low adhesion contributes to muscle and germ cell fates in P1 lineage, is a deep developmental biology and biophysical question worth future investigation (Figure 5b).

Down to the cellular scale, the ten cells with highest adhesion exhibit adhesion asymmetry between their daughter cells; however, whether the anterior or posterior daughter inherits the higher adhesion appears to be cell-dependent in a local manner (Figure 5b; Table S12). For 11583 cell measurements in 3 HMR-1 embryos, the cell adhesion surfaces were generally invariable with over 90% of cells having a variation level lower than 0.2 ratio (Figure 6a), indicating a high level of reproducibility of adhesion surfaces in developing embryos, which might drive anterior-posterior, left-right, and dorsal-ventral development (Figure 6, a and b). Time-lapse and cumulative cellular expressions of 36 sub-lineages are demonstrated in Table S12. Directional asymmetries, measured by *MolQuantifier* Module of daughter cell adhesion patterns may relate to differentiation (Figure 6c; Table S13). ABpl, ABaraa, and ABarap showed positive asymmetries (the anterior daughter cell showed higher adhesion related fluorescence than the posterior cell). Cp, Cpa and Cpp differentiate as skin and muscle (Table S14). These 2 daughter cells and their antecedents have high negative asymmetries and accurately move to form skin and muscle tissues, covering and supporting the entire embryo. The high asymmetric of fluorescent signals of ABalaa and ABalpp (which differentiate to neuronal and apoptotic tissues, respectively), and ABarpa and ABplaa (which differentiate to neuronal, skin, and death tissues) (Figure 6c). The ABarpaap-ABarpapp contact pair detected with the highest HMR-1/E-cadherin accumulation is further found to be a conserved contact among the 3 embryos (individuals), regardless if they are mechanically compressed or not. Whether this mechanically robust, conserved cell-cell contact with

substantially high adhesion serves a biological function is worth investigation, such as signaling transduction and cell division orientation, and positional control for further progenies that potentially interact with each other or form tissue/organ at specific positions.

Strikingly, quantification of both cellular (152 unique cells) and subcellular (478 unique cell–cell contact pairs) adhesion revealed exceptional precision across embryo samples, from the embryogenesis initiation to gastrulation, with a goodness of fit > 0.867 (proportional relationship) and an average correlation coefficient < 0.931 (Figure 6d; Table S15). In metazoan development, induction of cellular fates via Wnt and Notch signaling [55-57], often relies on continuous and direct contact and adhesion between specific cells to ensure proper functional interactions [4, 58, 59], for instance, interactions between ligands and their receptors to initiate signaling pathways [16, 17, 60]. By using digitized and detailed morphological data provided in this study, researchers can infer these cellular interactions by examining factors like the area and duration of cell contact. These cell-to-cell interactions and their adhesion level play a pivotal role in triggering cellular changes. The reproducibility of adhesion in cell-cell contacts was examined with 3 time-lapse samples involving 3613 contact pairs over 40008 time points with quantitative measurements of *HMR-1* fluorescence. This showed that a moderate level of variability was tolerated in cell-cell contacts, with 64% of data points having a variation ratio lower than 0.3 (Figure 6d). We demonstrated that cell-cell interaction adhesions controlled by remarkable invariances and reproducibility for precise cellular development, like cell volume, surface area, and contact area, in previous studies [7, 61, 62]. The 5 contact pairs with the highest *HMR-1* fluorescence over at least 3 consistent time points (Figure 6e), except for ABprpaaa, are skin and intestine cells, as with muscle cells, where over 80% of cells have high inferred adhesion levels. This might reflect the need to maintain a tighter, fixed shape. With the most *HMR-1* signals, during the mid-stage of the contact between ABarpaap and ABarpapp, a global rotational rearrangement of these 2 cells occurs, accompanied by pronounced deformation of ABarpaap; notably, the adhesion between the two cells reaches its maximum during this rotation, suggesting a genetically programmed enhancement of adhesion to preserve their relative position and morphology during development (Figure 6f).

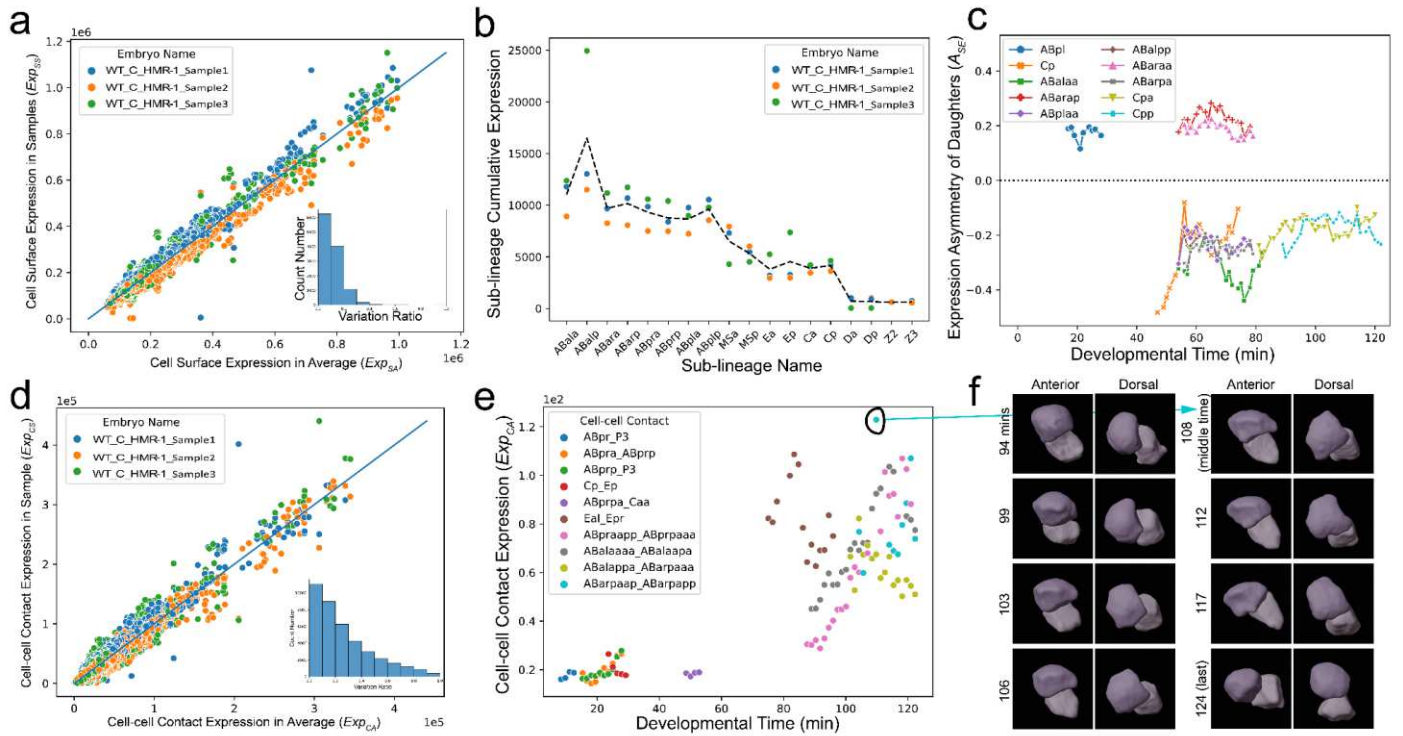


Fig 6. Whole-organism and subcellular-resolved quantification of *HMR-I/E*-cadherin dynamics governing adhesion. (a) Precision of cellular adhesion across their complete existing durations among three embryo samples. Inset: histogram of variation coefficients. (b) The cumulative expression of surface adhesion of the 18 spatial cell sub-lineages. (c) Cellular adhesion asymmetry between daughters of the ten cells with highest adhesion (presented in time-lapse curves), revealing persistent adhesion asymmetry with alternating directional biases. (d) Precision of subcellular adhesion across their complete existing durations among three embryo samples. Inset: histogram of variation coefficients. (e) Subcellular adhesion of the five contact pairs with the strongest adhesion and five contact pairs with the weakest (normalized absolute value), presented in time-lapse curves, reveals persistent, fluctuating, or trending adhesion over time. (f) Interacting cells with the largest expression contact area over their migration and morphogenesis: ABarpaap_ABarpapp.

Discussion

For about two decades, fluorescence labeling on different biological parts has been a widely used way to extract information, especially in the 3D + t scenario, where numerous data points are generated in a standard format. Typically, cell nucleus labeling allows cell lineage tracing considering its condensed signal and high contrast with respect to a black background, whereas cell membrane labeling enables complete cell morphology reconstruction. While cell nuclei cannot provide the complete information of membranes, the cell surface, as the complete cell body in space, may cover the needed information for cell lineage tracing that usually depends

on the cell nucleus, which is a point inside the cell shape. Inspired by this fact, this study devised an experimental-computational framework, *CTransformer*, to automatically reconstruct cell morphology (shape) and trace cell lineage with only one fluorescence channel for cell membrane labeling, bypassing laborious human intervention required before (Figure 1). Demonstrated with the worm *C. elegans*, *TUNETr* Module of *CTransformer* outperforms other state-of-the-art algorithms throughout embryogenesis up to a crowded stage with over 550 cells within its eggshell, in terms of both high segmentation accuracy (Figure 2) and low cell loss (Figure 3). Since the cell morphology reconstruction overcomes previous failure in cells with extremely deformed shapes (e.g. skin cells), and extremely small sizes (e.g. apoptotic cells), whose membrane signals were too weak and blurry, *m2nGAN* Module is able to make use of them for generating pseudo-nuclei images and cell lineage tracing (Figure 4). Such accurate cell morphology reconstruction reduces the potential for an expanding chain error in cell lineage tracing, jointly sparing the original cell nucleus channel so that *MolQuantifier* Module can use it to label another molecule of interest, exemplified by the well-known and critical adhesive protein *HMR-1*/E-Cadherin (Figure 5). Quantitative labeling demonstrates previously unknown features of precisely regulated adhesion of specific cells and cell-cell contacts (Figure 6).

By full utilization of image information and enabling subcellular-resolution molecular quantification, *CTransformer* is reasonably applied for a broad range of scenarios.

1. The spared channel(s) can accommodate labeling not only on the *HMR-1*/E-Cadherin, but also any other labelable molecule. When modulating fluorescence colors and applying three or more color markers, the dynamics of two or more molecules can be jointly quantified. This permits landscape scanning beyond the previous dimensions, including but not limited to gene expression, chromatin accessibility. And data with different dimensions can be studied as a whole.
2. Cell properties such as adhesion (labeled by *HMR-1*/E-Cadherin), tension (labeled by F-actin), and others can be estimated without harming the sample. Compared with other unarmful indirect estimation methods represented by image-based mechanical modeling/calculation [63, 64], *CTransformer* set up the focus on molecular labeling, which is closer to the reality since the force interpreted is directly attributed to a certain molecule without ambiguity. Conversely, indirect estimation methods, even if the inferred force is correct, represent a resultant effect and miss the detailed information, such as the origins of this force.
3. The general design of *CTransformer* unleashes its applicability not only on the exemplary *C. elegans* embryo, but also on other systems with a comparable level of cell crowdedness. For example, the Ascidian system is traced with its first half of embryogenesis, aided by a large amount of manual tracing. *CTransformer* could be applied in such a scenario, and also the systems with abnormal conditions (RNAi, mutation, etc.), as long as there is a cell membrane fluorescence channel. Importantly, the reliance on only cell membrane fluorescence but not nucleus fluorescence means that the 3D + t

imaging can have roughly half of photobleaching and phototoxicity, and equivalently, allowing twice the duration for the sample to image.

The whole *CTransformer* framework is applied to both previously published and newly added *C. elegans* embryos, providing an informative resource including 47 embryos for both developmental biology and computer science research. Dated back to the developmental path of cell nucleus tracing of StarryNite/AceTree to gene expression profiling of AceBatch, followed by the development path of cell morphology reconstruction algorithms from 3DMMS, *CShaper*, *CMap*, and *CTransformer*, the field is transitioning toward an advanced era of measuring molecular distribution on top of the previous both.

Methods

***C. elegans* strains, molecular probes and imaging specifications for cell adhesion quantification**

All animals were maintained on nematode growth media (NGM) plates seeded with *Escherichia coli* (OP50) at room temperature. First, for embryos subjected to nucleus-based cell lineage tracing and membrane-based cell morphology segmentation (uncompressed), strains ZZY0868, ZZY0861 and ZZY0655 were used [4, 7]. The transgenic strain RB669 (*wee-1.1(ok418)*) was crossed with the strain ZZY0861 [4], resulting in a new strain ZZY1142 with both *wee-1.1* deficiency and fluorescence labeling on the cell nucleus and membrane. Second, a cell-membrane-labeled strain LP172 and ZZY0637 (carrying a single copy of PH2::mCherry) [7], was crossed with the HMR-1-labeled strain LP172 (carrying a GFP-tagged endogenous *HMR-1* gene) [47]. Both markers were rendered homozygous in the resulting strain ZZY0868 (*hmr-1(cp21[hmr-1::GFP + LoxP])* I; *unc-119(tm4063)* III; *zzySi139 [Phis-72::PH(PLC1delta1)::mCherry::pie-1 3' UTR + unc-119(+)]* II).

Embryos expressing both reporters were imaged live on a Leica SP5 laser-scanning confocal microscope, following the established miniMos-based method [7]. Sequential high-resolution z-stacks (68 slices per stack with a z-resolution of 0.42 μm) were acquired at intervals of approximately 1.5 minutes, using 488 nm and 561 nm laser lines for excitation of the HMR-1::GFP and PLC1 δ 1::mCherry signals, respectively. Z-axis compensation (19-95% for the 561 nm laser) was applied to mitigate signal attenuation. This imaging regimen captured the dynamic localization of adhesion complexes relative to cell morphology throughout early embryogenesis. The resultant four-dimensional data were exported as multi-TIFF stacks for subsequent processing and analysis by the *MolQuantifier* Module.

3D + t fluorescence imaging and confocal details

For *C. elegans* embryos with both fluorescence labeling on the cell nucleus (GFP) and membrane (mCherry), the imaging protocol followed the previous studies [7, 65]. Fluorescence images start to be captured in embryos with 4 cells at most using either the SP5 II or Stellaris confocal microscopy system from Leica. Then, they are processed and analyzed using the LAS X software from Leica and ImageJ. To keep the embryos alive, the intensity of light was set to be low (a scanning speed of 8,000 Hz with a water immersion objective). The imaging process was separated into successive time blocks of 60 time points. The z -axis compensation ranged from 0.5-3% for the 488-nm laser and 20-95% for the 594-nm laser. Subsequently, *StarryNite* and *AceTree* [36, 66, 67], were used to recognize every cell nucleus (cell index: i) at consecutive time points (time index: t) in each individual embryo (embryo index: j), systematically recording their 3D positions $(x_{i,j,t}, y_{i,j,t}, z_{i,j,t})$.

For the embryos with both fluorescence labeling on HMR-1 and cell membrane (mCherry), the imaging protocol followed a combination of methods described in our previous studies [7, 65], as detailed here. One to four-celled embryos were mounted for imaging using 1% methylcellulose in Boyd's buffer with 20 μm Polybead® microspheres (Polysciences, Inc.); fluorescence imaging was performed with an inverted Leica SP5 confocal microscope equipped with two hybrid detectors at a constant ambient temperature of 21 °C. Images were simultaneously collected for both GFP and mCherry channels using a water immersion objective; both channels were imaged with a frame size of 712×512 pixels (0.09 $\mu\text{m}/\text{pixel}$) combined with a scanning speed of 8,000 Hz using a resonance scanner. The excitation laser beams used for GFP and mCherry were 488 nm and 594 nm, respectively; images containing 68 z -steps were collected for three individual embryos with a z -resolution of 0.42 μm from top to bottom for every time point, which was at 1.41-minute intervals. Images were continuously collected for 100 time points with the embryos developing into “E4” (“E”, the intestine progenitor, dividing into four granddaughters) or “E8” stages (“E”, the intestine progenitor, dividing into four granddaughters). The z -axis compensation was 0.5-8% for the 488-nm laser and 19-95% for the 594-nm laser with the pinhole size of 2.5 AU (airy unit). Later, a deconvolution process was used for fluorescence images before the application of *CTransformer* framework. During imaging, the temporal resolution was 1.39 to 1.45 min per frame, and the volumetric spatial resolution was 0.18 or 0.25 (0.18^3 or $0.25^3 \mu\text{m}^3$ per voxel).

Detailed Training and Running Segmentation Workflow

The running steps of *TUNETr*, with real nuclei fluorescence images, conclude recognizing membranes and generating nuclei-prompting seeds (Figure S1). The comparison of *CTransformer* is also conducted with the results under *sTUNETr* and nuclei-prompting strategy (Figures 2 and 3).

Step 1: Recognizing Cell Membrane with Transformer-based Regression Large Model and Topology-constraint Loss. In the initial phase of our method, as outlined in (step 1 of Figure S2a), we establish a ground truth (GT) dataset comprising manually annotated images, like other DNNs. These annotations represent a range of embryonic developmental stages and imaging conditions. This GT dataset serves as the foundation for training the *TUNETr* within *CTransformer*, yielding a pretrained *iTUNETr* network with 71 volumes in this study. We utilize the capability of this large-scale model, convoluting voxels of 3D volumes with their neighboring voxels (self-attention). So, *iTUNETr* facilitates the precise identification of voxels that were previously indistinguishable due to extremely weak signals, noises, and various artifacts (Figure 3a). Consequently, the recognition results in seen conditions (similar raw images in the GT) are granted by the DNN (*iTUNETr*). With pretrained *iTUNETr*, we introduce a topology loss term L_t (step 1 on the right of Figure S1A), in further training. Pretrained *iTUNETr* is necessary since the topology loss can only be accurately computed with the context of a relatively correct spatial shape (cellular). This improved loss function is an enhancement of the geometric loss term L_g (step 1 on the right of Figure S2A), being effective to constrain a single cell's topological structure in the same way as the GT. This represents a minor adjustment to the initial model *iTUNETr*, from a higher-dimensional perspective (topological latent space). There are two ways to annotate the membrane for recognition. One approach starts from scratch, relying solely on human visual inspection to annotate images on each 2D slice and then observing in 3D to check if the annotations across different slices are continuous and credible. These annotations are voxel-wise accurate, but are subject to variability due to different annotators' standards and encounter issues with discontinuity in 3D. However, this can lead to issues where pre-recognized volumes influence human judgment, as some areas in the raw volumes are not absolutely clear, even with human annotation. So, we use both approaches to manually annotate the GT for training, to discard the dross and select the essence. Here, we get the "specialist" DNN, *iTUNETr*.

Steps 2 and 3: Synthetic Ground Truth of 4D Fluorescence Images and Fine-tuning Semi-supervised Learning Strategy. We trained the "specialist" DNN, using manually annotated GT, to recognize the membrane (semantic segmentation) for fluorescence 3D images. To overcome the biggest drawback of DNN, lacking generalization ability and only working on a specific task, and also for transformer-based large models, we make a substantial synthetic GT and semi-supervised learning scheme that cooperates with human intervention. The *iTUNETr* model recognizes cell membranes in anisotropic raw fluorescence images, achieving an accuracy comparable to human annotation level for seen imaging conditions (with a Dice Score larger than 0.9). To amass a large GT dataset while maintaining a human level of accuracy, *TUNETr* allows users to manually adjust the nuclei promptings via StarryNite and AceTree [36, 66, 67]. This human-in-the-loop step provides location prompts and eliminates irrelevant surroundings from the recognition area,

refining the membrane recognition to reasonable positions and sizes. Users can repeatedly tune the prompts and select the correct membrane recognition volumes to form the synthetic GT, integrating it with the manually annotated GT (first row of Figure 1b and step 2 of Figure S2a). These new GT membrane annotations are constrained from real cell shape, and the embryos' edges are tripled in thickness, to draw DNN's attention to the weak imaging area. This user-customized synthetic GT encompasses a vast amount of unlabeled data, covering all imaging conditions (both seen and unseen, labeled and unlabeled) relative to the task. Then, restoring from *iTUNETr*, adjusting the DNN parameters, and proceeding with training (step 3 of Figure S2a), *TUNETr* generates a "generalist" DNN, *sTUNETr* (semi-supervised *TUNETr*, akin to *iTUNETr*), tailored for specialized tasks under unseen imaging conditions, such as those with significant artifacts or in late developmental stages.

Step 4: Generate Reliable Nuclei-prompting Seeds with Well-trained DNN. We introduce an additional nuclei-prompting step to keep the chain and serious cell shape segmentation errors away (step 4 of Figure S2a). Informed by nucleus promptings, *TUNETr* avoided completely erroneous segmentation in extremely ambiguous imaging areas. This step helps in constructing potent cell instance generation (multi-cell segmentation) algorithms from cell membrane recognition. Due to the DNN model is an image-wise global and customized method, the recognition of previous studies for top and bottom areas (axially upper and lower) in the volumetric images still failed when the imaging conditions are very different, or the trained model has not "met" before [7, 33, 68] (in the training data). Fortunately, nucleus images are relatively clear and easy to detect as nucleus instances.

Steps 5 to 7: Producing Multiple Single-cell Segmentation for 4D Fluorescence Images. At the final processing steps (steps 5 to 7 of Fig. S2a), we run the segmentation on 34 embryos, providing a dataset for fluorescence images processing and relative biomedical studies. Integrating the large training set and *sTUNETr*, we generate a more convincing cell morphology map. In this work, we provide interface software and a corresponding website for this 4D cellular segmentation map, also compatible with the visualization database website (CMOS, bcc.ee.cityu.edu.hk/cmos) and analysis software (ITK-SNAP-CVE) [4, 7]. While running multiple cell segmentation scripts, the program is relatively friendly to non-programmers, biologists, and physicists. Therefore, for the different multicellular worm *C. elegans*, *TUNETr* Module can directly generate 4D data that can be traced back to the cell shape level, which can greatly promote the study of cell morphology in the embryonic development process.

4dTUNETr

We improved the input and output of the DNN of *TUNETr* (Figures S1 and S2, see Supplementary Information), as well as the cell instance segmentation for a better and more robust single-cell segmentation for pseudo nuclei images generation. We give clear nuclei sizes and positional hints (both from segmented single-cell morphology) for GAN to generate the nuclei with relatively accurate nuclei fluorescent signals (Figure S3). Thus, *TUNETr* was trained with only the membrane channel (*iTUNETr* and *sTUNETr* used the nucleus channel for training and running). To generate stable recognition of membranes, the 4D images from the synthetic dataset were used for training in semi-supervised learning *sTUNETr* (section Detailed Training and Running Segmentation Workflow). For the 3 input sequential volumes, $T-1$, T , $T+1$, the output would be predictions of these 3 time points. $T-1$ and $T+1$ would also generate $T-2$, $T-1$, T , and, T , $T+1$, $T+2$, respectively. As for a shifted window, at time point T , *4dTUNETr* produced 3 membrane predictions T' , T'' , and T''' . The binarization process takes advantage of these predictions to integrate a robust membrane segmentation. In one of the evaluated embryos, *4dTUNETr* produces 524 while *sTUNETr* produces 506 cells for 550-cell embryos.

m2nGAN: The GAN and Generative Network for Fluorescent Pseudo Nuclei Images

Segmented single-cell results from *4dTUNETr* (improved *sTUNETr*) were fed to the 3D Cycle Generative adversarial network (3D-cycle-GAN) to generate time-lapse pseudo-nuclei fluorescence images [69, 70]. Only membrane-labelled 4D images (*4dTUNETr*) are needed to build the cell-resolved morphology map for living organisms.

With single-cell segmentations from *4dTUNETr*, an octahedron, is artificially generated for every cell as hints for *m2nGAN*. These octahedra are produced by dilation operations with N iterations where N is derived from the R of one cell, formulated as

$$N = \min[n \in \mathbb{Z} \mid x \leq R].$$

And the R is calculated as

$$R = \sqrt[3]{\frac{3 \times V_{\text{voxel}}}{4 \times \pi}}.$$

The voxels of localized and size-fixed octahedra were set as 255 in the 8-bit volumetric images. With multiple octahedra in a 3D volume, *m2nGAN* can transform the distributions of images to the real nuclei fluorescence images (Figure S3).

MolQuantifier

We develop the *MolQuantifier* Module simultaneously to quantify the modular distributions on every single cell. *MolQuantifier* could measure the RNA expression imaged in the fluorescence images. On every segmented cell surface (it consists of three layers of voxels, simulating the real structure of cell membranes), *MolQuantifier* calculates all the expression values on the membrane. This quantification method can accurately obtain expression in the 3D selected area, like the membrane, nucleus, cytoplasm, and contacting surface, rather than only in the estimated whole cell area.

Data and code availability

Practical and efficient scalable software is provided for time-lapse single-cell volumetric analysis, and 4D embryonic segmentation from fluorescent images of embryos. The code, software, raw and segmented data are available in link <https://doi.org/10.6084/m9.figshare.27085657>. More source data will be provided upon reasonable request.

Acknowledgements

We thank Prof. Xiaojing Yang (Peking University) and Fengxi Yu (Tsinghua University) for their invaluable assistance in initiating this project. This work is supported by Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA) and Hong Kong Research Grants Council (Project 11204821) to H.Y., and by Hong Kong Innovation and Technology Commission (GHP/176/21SZ) and Hong Kong Research Grants Council (Projects HKBU12101520, HKBU12101522, HKBU12101323) to Z.Z.

References

1. Fujii, Y., et al., *Spatiotemporal dynamics of single cell stiffness in the early developing ascidian chordate embryo*. Communications Biology, 2021. **4**(1).
2. Brangwynne, C.P., et al., *Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation*. Science, 2009. **324**(5935).
3. Otsuji, M., et al., *A mass conserved reaction-diffusion system captures properties of cell polarity - PubMed*. PLoS computational biology, 2007. **3**(6).
4. Guan, G., et al., *Cell lineage-resolved embryonic morphological map reveals signaling associated with cell fate and size asymmetry*. Nature Communications, 2025. **16**(1).
5. Stegmaier, J., et al., *Real-Time Three-Dimensional Cell Segmentation in Large-Scale Microscopy Data of Developing Embryos*. Developmental Cell, 2016. **36**(2): p. 225-240.
6. Azuma, Y. and S. Onami, *Biologically constrained optimization based cell membrane segmentation in C. elegans embryos*. BMC Bioinformatics, 2017. **18**(1): p. 307.
7. Cao, J., et al., *Establishment of a morphological atlas of the Caenorhabditis elegans embryo using deep-learning-based 4D segmentation*. Nature Communications, 2020. **11**(1): p. 6254.
8. Wang, F., et al., *Far-field super-resolution ghost imaging with a deep neural network constraint*. Light: Science & Applications, 2022. **11**(1).

9. Guignard, L., et al., *Contact area-dependent cell communication and the morphological invariance of ascidian embryogenesis*. Science, 2020. **369**(6500): p. eaar5663.
10. Waliman, M., et al., *Automated cell lineage reconstruction using label-free 4D microscopy*. GENETICS, 2024. **228**(2).
11. Cao, J., et al., *3DMMS: robust 3D Membrane Morphological Segmentation of C. elegans embryo*. BMC Bioinformatics, 2019. **20**(1).
12. Cao, J., et al., *CShaperApp: Segmenting and analyzing cellular morphologies of the developing Caenorhabditis elegans embryo*. Quantitative Biology, 2024.
13. Qiu, C., et al., *A single-cell time-lapse of mouse prenatal development from gastrula to birth*. Nature 2024 626:8001, 2024. **626**(8001): p. 1084-1093.
14. Heymann, J.A.W., et al., *Site-specific 3D imaging of cells and tissues with a dual beam microscope*. Journal of Structural Biology, 2006. **155**(1): p. 63-73.
15. Abouakil, F., et al., *An adaptive microscope for the imaging of biological surfaces*. Light: Science & Applications, 2021. **10**(1).
16. Chen, L., et al., *Establishment of Signaling Interactions with Cellular Resolution for Every Cell Cycle of Embryogenesis*. Genetics, 2018. **209**(1): p. 37-49.
17. Ma, X., et al., *A 4D single-cell protein atlas of transcription factors delineates spatiotemporal patterning during embryogenesis*. Nature Methods 2021 18:8, 2021. **18**(8).
18. Ho, V.W.S., et al., *Systems-level quantification of division timing reveals a common genetic architecture controlling asynchrony and fate asymmetry*. Molecular Systems Biology, 2015. **11**(6): p. 814.
19. Maître, J.-L. and C.-P. Heisenberg, *Three Functions of Cadherins in Cell Adhesion*. Current Biology, 2013. **23**(14): p. R626-R633.
20. Calcutt, R., et al., *Plant cell adhesion and growth on artificial fibrous scaffolds as an in vitro model for plant development*. Science Advances, 2021. **7**(43).
21. Vaswani, A., et al., *Attention is all you need*. Advances in neural information processing systems, 2017. **30**.
22. Panaretos, V.M. and Y. Zemel, *Statistical Aspects of Wasserstein Distances*. Annual Review of Statistics and Its Application, 2019. **6**(1): p. 405-431.
23. Clough, J.R., et al., *A Topological Loss Function for Deep-Learning Based Image Segmentation Using Persistent Homology*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022. **44**(12): p. 8766-8778.
24. Yushkevich, P.A., Y. Gao, and G. Gerig. *ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images*. in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2016. IEEE.
25. Moore, J.L., Z. Du, and Z. Bao, *Systematic quantification of developmental phenotypes at single-cell resolution during embryogenesis*. Development, 2013. **140**(15): p. 3266-3274.
26. Du, Z., et al., *De novo inference of systems-level mechanistic models of development from live-imaging-based phenotype analysis*. Cell, 2014. **156**(1): p. 359-372.
27. Fan, D., et al., *Spatiotemporal analysis of mRNA-protein relationships enhances transcriptome-based developmental inference*. Cell Reports, 2024. **43**(3).
28. Cole, A.G., et al., *Gene regulatory patterning codes in early cell fate specification of the C. elegans embryo*. Elife, 2024. **12**: p. RP87099.
29. Hatamizadeh, A., et al., *Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images*, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. 2022, Springer International Publishing. p. 272-284.
30. He, Y., et al., *SwinUNETR-V2: Stronger Swin Transformers with Stagewise Convolutions for 3D Medical Image Segmentation*, in *Lecture Notes in Computer Science*. 2023, Springer Nature Switzerland. p. 416-426.
31. Zhou, Z., et al., *UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation*. IEEE Transactions on Medical Imaging, 2020. **39**(6): p. 1856-1867.
32. Milletari, F., N. Navab, and S.-A. Ahmadi. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. in *2016 Fourth International Conference on 3D Vision (3DV)*. 2016. IEEE.
33. Eschweiler, D., R.S. Smith, and J. Stegmaier. *Robust 3d Cell Segmentation: Extending The View Of Cellpose*. in *2022 IEEE International Conference on Image Processing (ICIP)*. 2022. IEEE.
34. Weigert, M., et al. *Star-convex polyhedra for 3D object detection and segmentation in microscopy*. in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2020.
35. Chisholm, A.D. and J. Hardin, *Epidermal morphogenesis*. 2005: *WormBook*.

36. Bao, Z., et al., *Automated cell lineage tracing in Caenorhabditis elegans*. Proceedings of the National Academy of Sciences, 2006. **103**(8): p. 2707-2712.
37. Santella, A., et al., *A hybrid blob-slice model for accurate and efficient detection of fluorescence labeled nuclei in 3D*. BMC Bioinformatics, 2010. **11**(1): p. 580.
38. Santella, A., Z. Du, and Z. Bao, *A semi-local neighborhood-based framework for probabilistic cell lineage tracing*. BMC Bioinformatics, 2014. **15**(1): p. 217.
39. Li, Z., et al., *Deep learning-based enhancement of fluorescence labeling for accurate cell lineage tracing during embryogenesis*. Bioinformatics, 2024. **40**(11).
40. Zhao, Z., et al., *Single-cell dynamics of chromatin activity during cell lineage differentiation in Caenorhabditis elegans embryos*. Molecular Systems Biology, 2021. **17**(4): p. e10075.
41. Xu, W., et al., *A lineage-resolved cartography of microRNA promoter activity in C. elegans empowers multidimensional developmental analysis*. Nature communications, 2024. **15**(1): p. 2783.
42. Del Rosario, M., et al., *PhotoFiTT: a quantitative framework for assessing phototoxicity in live-cell microscopy experiments*. Nature Communications, 2025.
43. Icha, J., et al., *Phototoxicity in live fluorescence microscopy, and how to avoid it*. BioEssays, 2017. **39**(8): p. 1700003.
44. Rasmussen, J.P., et al., *Cell Interactions and Patterned Intercalations Shape and Link Epithelial Tubes in C. elegans*. PLoS Genetics, 2013. **9**(9): p. e1003772.
45. Broadbent, I.D. and J. Pettitt, *The C. elegans hmr-1 gene can encode a neuronal classic cadherin involved in the regulation of axon fasciculation*. Current Biology, 2002. **12**(1): p. 59-63.
46. Halbleib, J.M. and W.J. Nelson, *Cadherins in development: cell adhesion, sorting, and tissue morphogenesis*. Genes & development, 2006. **20**(23): p. 3199-3214.
47. Marston, D.J., et al., *MRCK-1 drives apical constriction in C. elegans by linking developmental patterning to force generation*. Current Biology, 2016. **26**(16): p. 2079-2089.
48. Schnabel, R., et al., *Assessing normal embryogenesis in Caenorhabditis elegans using a 4d microscope: Variability of development and regional specification*. Developmental biology, 1997. **184**(2): p. 234-265.
49. Fickentscher, R., S.W. Krauss, and M. Weiss, *Anti-correlation of cell volumes and cell-cycle times during the embryogenesis of a simple model organism*. New Journal of Physics, 2018. **20**(11): p. 113001.
50. Azuma, Y., H. Okada, and S. Onami, *Systematic analysis of cell morphodynamics in C. elegans early embryogenesis*. Frontiers in Bioinformatics, 2023. **3**: p. 1082531.
51. Van Bavel, C., W. Thiels, and R. Jelier, *Cell shape characterization, alignment, and comparison using FlowShape*. Bioinformatics, 2023. **39**(6).
52. Murray, J.I. and Z. Bao, *Automated lineage and expression profiling in live Caenorhabditis elegans embryos*. Cold Spring Harbor protocols, 2012. **2012**(8): p. pdb. prot070615.
53. Packer, J.S., et al., *A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution*. Science, 2019. **365**(6459): p. eaax1971.
54. van der Burght, S.N., et al., *A lineage-resolved multimodal single-cell atlas reveals the genomic dynamics of early C. elegans development*. bioRxiv, 2024: p. 2024.12. 02.626321.
55. Walston, T., et al., *Multiple Wnt Signaling Pathways Converge to Orient the Mitotic Spindle in Early C. elegans Embryos*. Developmental Cell, 2004. **7**(6): p. 831-841.
56. Thorpe, C.J., et al., *Wnt Signaling Polarizes an Early C. elegans Blastomere to Distinguish Endoderm from Mesoderm*. Cell, 1997. **90**(4).
57. Neves, A. and J.R. Priess, *The REF-1 Family of bHLH Transcription Factors Pattern C. elegans Embryos through Notch-Dependent and Notch-Independent Pathways*. Developmental Cell, 2005. **8**(6): p. 867-879.
58. Pocock, R., et al., *Functional dissection of the C. elegans cell adhesion molecule SAX-7, a homologue of human LI*. Molecular and Cellular Neuroscience, 2008. **37**(1): p. 56-68.
59. Asan, A., S.A. Raiders, and J.R. Priess, *Morphogenesis of the C. elegans Intestine Involves Axon Guidance Genes*. PLOS Genetics, 2016. **12**(4): p. e1006077.
60. Du, Z., et al., *The Regulatory Landscape of Lineage Differentiation in a Metazoan Embryo*. Developmental Cell, 2015. **34**(5): p. 592-607.
61. Li, Z., et al., *An effective method for quantification, visualization, and analysis of 3D cell shape during early embryogenesis*. Quantitative Biology, 2025. **13**(1).
62. Guan, G., et al., *EmbSAM: cell boundary localization and Segment Anything Model for fast images of developing embryos*. Communications Biology, 2025.

63. Vanslambrouck, M., et al., *Image-based force inference by biomechanical simulation*. PLOS Computational Biology, 2024. **20**(12): p. e1012629.
64. Borghi, N., et al., *E-cadherin is under constitutive actomyosin-generated tension that is increased at cell–cell contacts upon externally applied stretch*. Proceedings of the National Academy of Sciences, 2012. **109**(31): p. 12568-12573.
65. Kuang, X., et al., *Computable early Caenorhabditis elegans embryo with a phase field model*. PLOS Computational Biology, 2022. **18**(1): p. e1009755.
66. Murray, J.I., et al., *Automated analysis of embryonic gene expression with cellular resolution in C. elegans*. Nature Methods, 2008. **5**(8): p. 703-709.
67. Boyle, T.J., et al., *AceTree: a tool for visual analysis of Caenorhabditis elegans embryogenesis*. BMC Bioinformatics 2006 7:1, 2006. **7**(1).
68. Wang, A., et al., *A novel deep learning-based 3D cell segmentation framework for future image-based disease detection*. Scientific Reports, 2022. **12**(1).
69. Ge, Y., et al. *Unpaired Mr to CT Synthesis with Explicit Structural Constrained Adversarial Learning*. in *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*. 2019. IEEE.
70. Isola, P., et al., *Image-to-image translation with conditional adversarial networks*, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. p. 1125-1134.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformationsubmitting1.pdf](#)