Machine Learning-Accelerated Analysis of *In Utero* Embryo Phenotyping in *C. elegans* for Reproductive Toxicity Assessment

Abhishri Medewar[1], Andrew DuPlissis[1], Adam Laing[1], Amber Shen[1], Evan Hegarty[1], Sebastian Gomez[1], Gina Carrion[1], Julia Brown[1], Sudip Mondal[1], and Adela Ben-Yakar[1,2,3]

[1]vivoVerse, LLC, Austin, TX 78731, USA
[2]Department of Mechanical Engineering, The University of Texas at Austin, TX 78712, USA
[3]Department of Biomedical Engineering, The University of Texas at Austin, TX 78712, USA

Corresponding Authors: sudip.mondal@vivoverse.com; ben-yakar@mail.utexas.edu
Primary Contact: Adela Ben-Yakar (ben-yakar@mail.utexas.edu)

Supplementary Information

Supplementary Table 1. Model performance for embryo model ablation experiments.

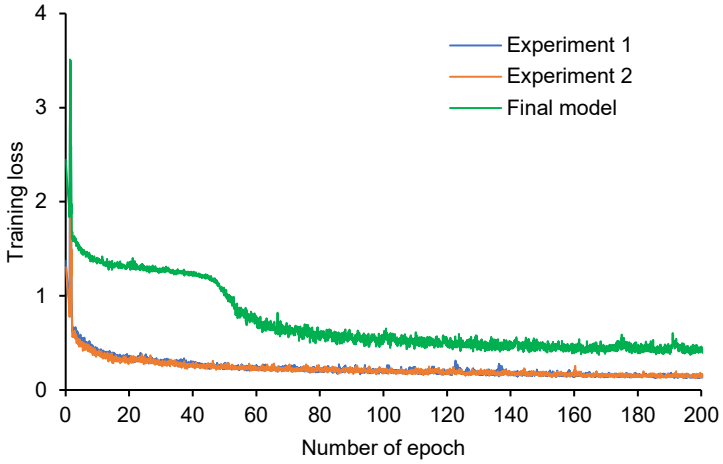
Supplementary Table 2. Statistical tests for well-level averages for 4 DMSO conditions.

Supplementary Table 3. Statistical tests for aggregated averages from 3-well combinations.

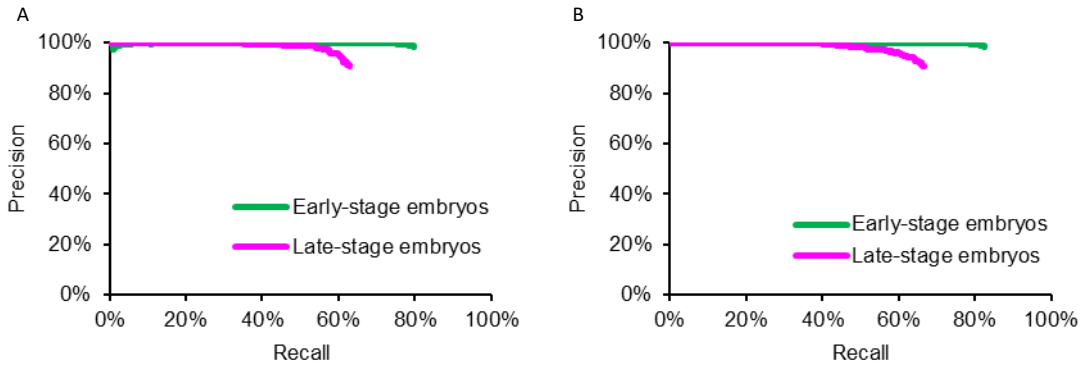
Supplementary Table 4. Statistical tests for aggregated CV% from 3-well combinations.

Supplementary Table 5. Effective concentration values for embryo-related parameters for propiconazole-treated worms.

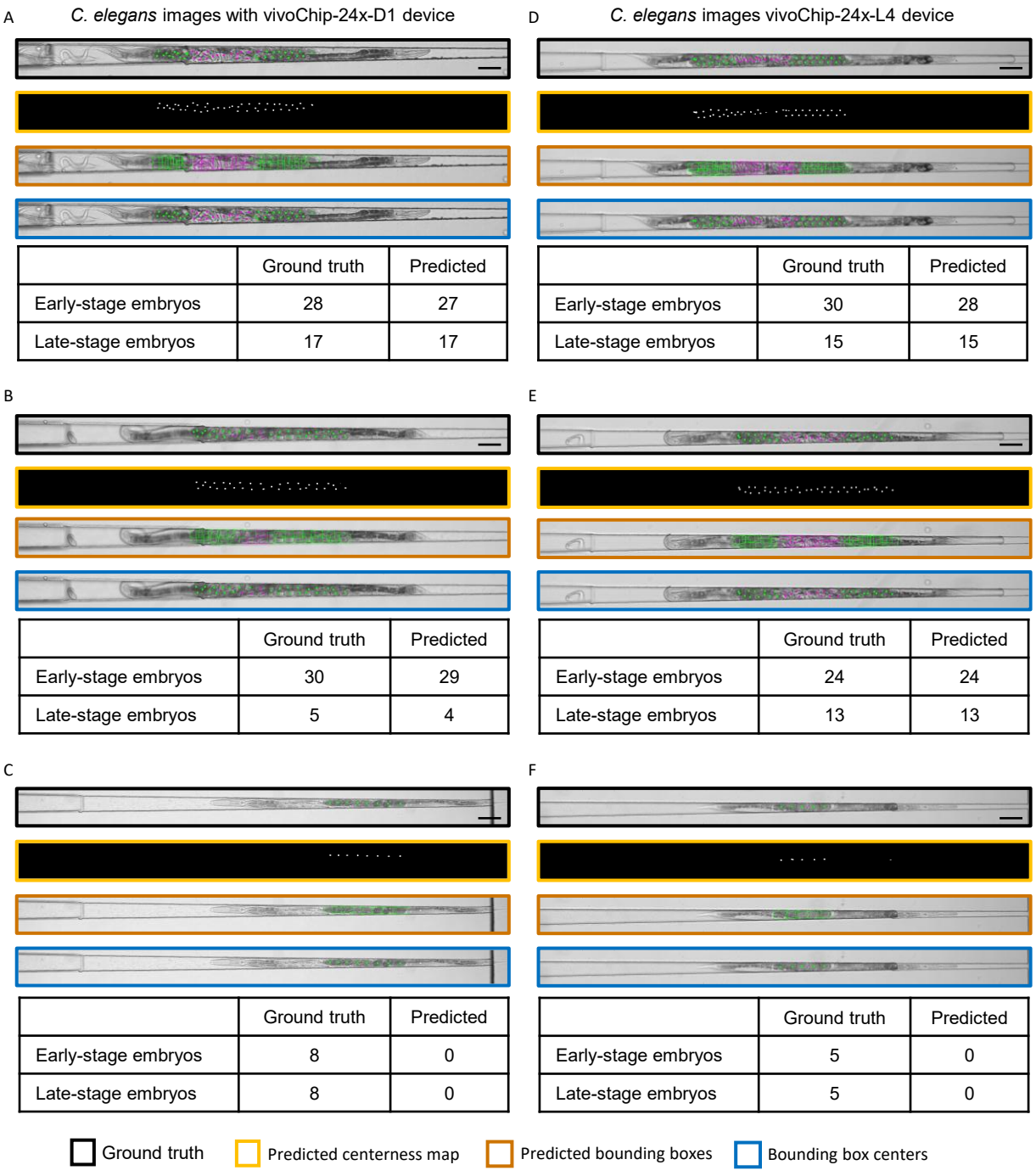
Supplementary Table 6. Effective concentration values for embryo-related parameters for methylmercury-treated worms.



Supplementary Figure 1. Training loss plots for three embryo models. The curve represents training loss for the first 200 epochs for embryo ablation experiment 1, ablation experiment 2, and final embryo model (EmbryoMAE-Det).



Supplementary Figure 2. Precision-recall curve for embryo ablation experiments. (A) Precision vs recall plot for early-stage and late-stage embryos for first ablation experiment (Experiment 1). **(B)** Precision vs recall plot for early-stage and late-stage embryos for second ablation experiment (Experiment 2).



	Ground truth	Predicted
Early-stage embryos	28	27
Late-stage embryos	17	17

	Ground truth	Predicted
Early-stage embryos	30	28
Late-stage embryos	15	15

	Ground truth	Predicted
Early-stage embryos	30	29
Late-stage embryos	5	4

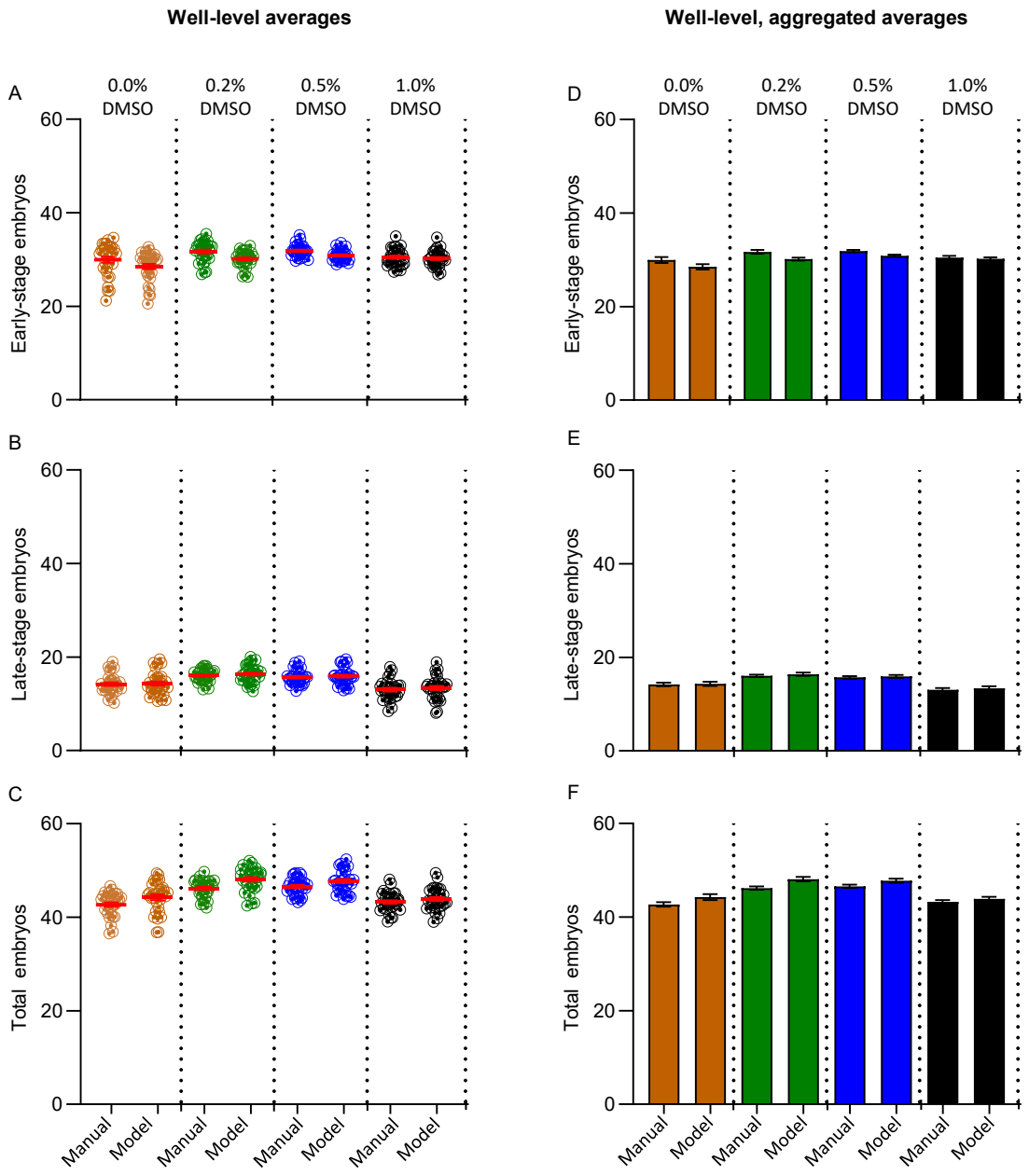
	Ground truth	Predicted
Early-stage embryos	24	24
Late-stage embryos	13	13

	Ground truth	Predicted
Early-stage embryos	8	0
Late-stage embryos	8	0

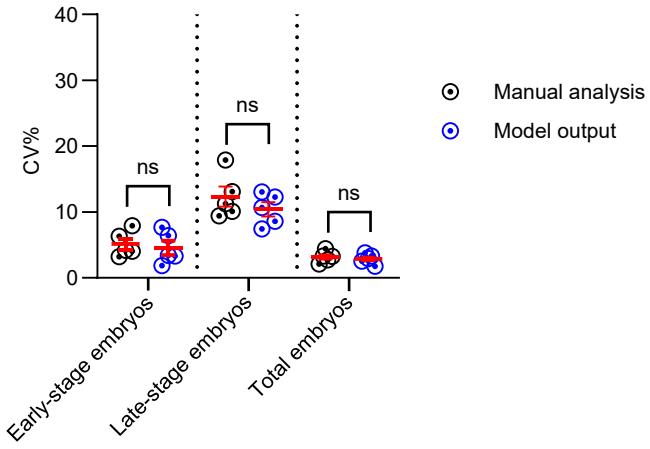
	Ground truth	Predicted
Early-stage embryos	5	0
Late-stage embryos	5	0

Supplementary Figure 3. Examples of *C. elegans* embryo detection using EmbryoMAE-Det model. Example images from test sets were used to compare the predicted embryo information with the ground truth data. **(A-C)** *C. elegans* captured in vivoChip-24x-D1 devices. **(D-F)** *C. elegans* captured in vivoChip-24x-L4 devices. The image shows the ground truth data, predicted centerness map, predicted bounding boxes, and the bounding box centers. The table summarizes the number of early-stage (green boxes and

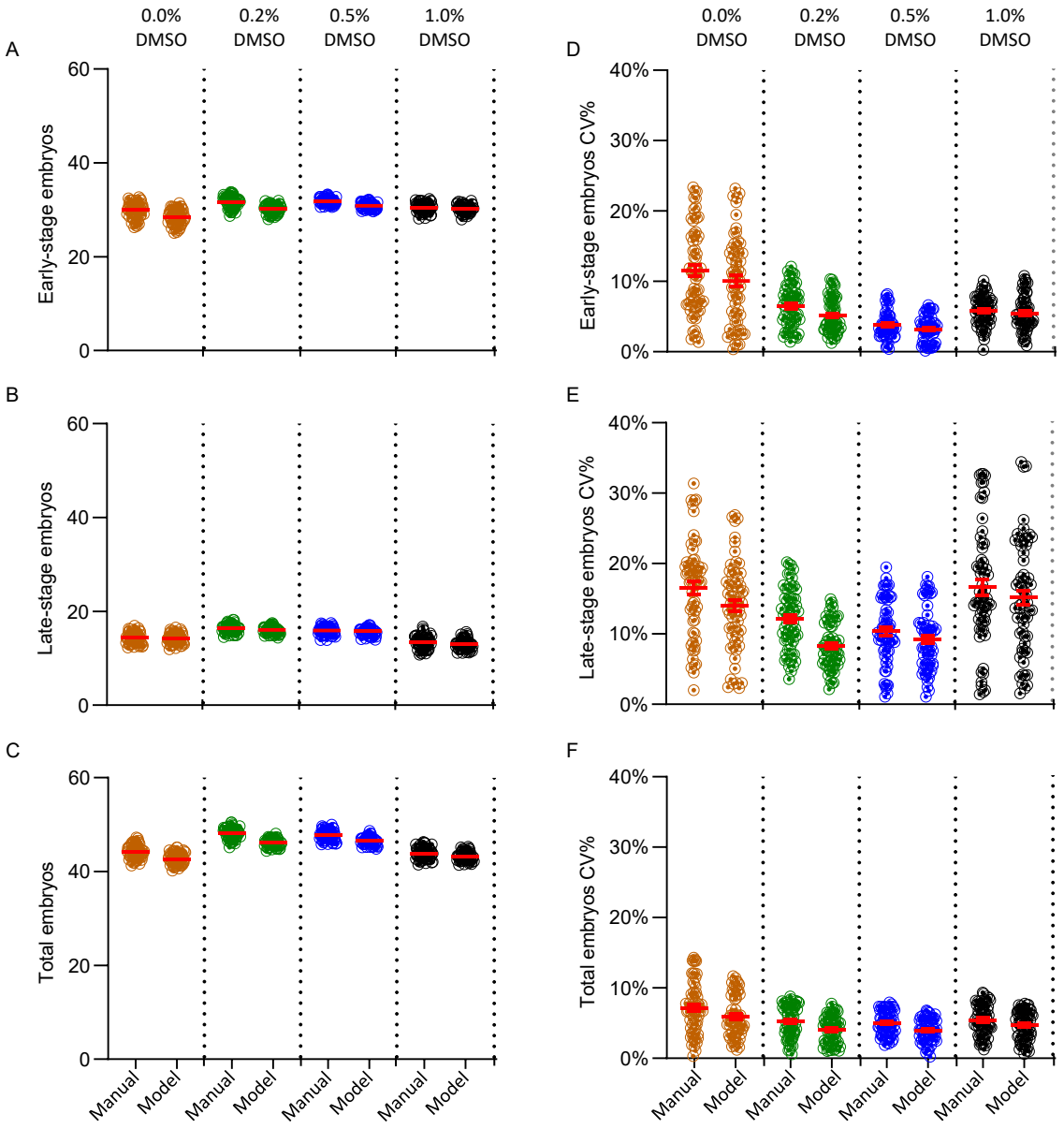
dots) and late-stage embryo (magenta boxes and dots) numbers for each worm, predicted by the model and manually scored. The scale bar = 200 μm.



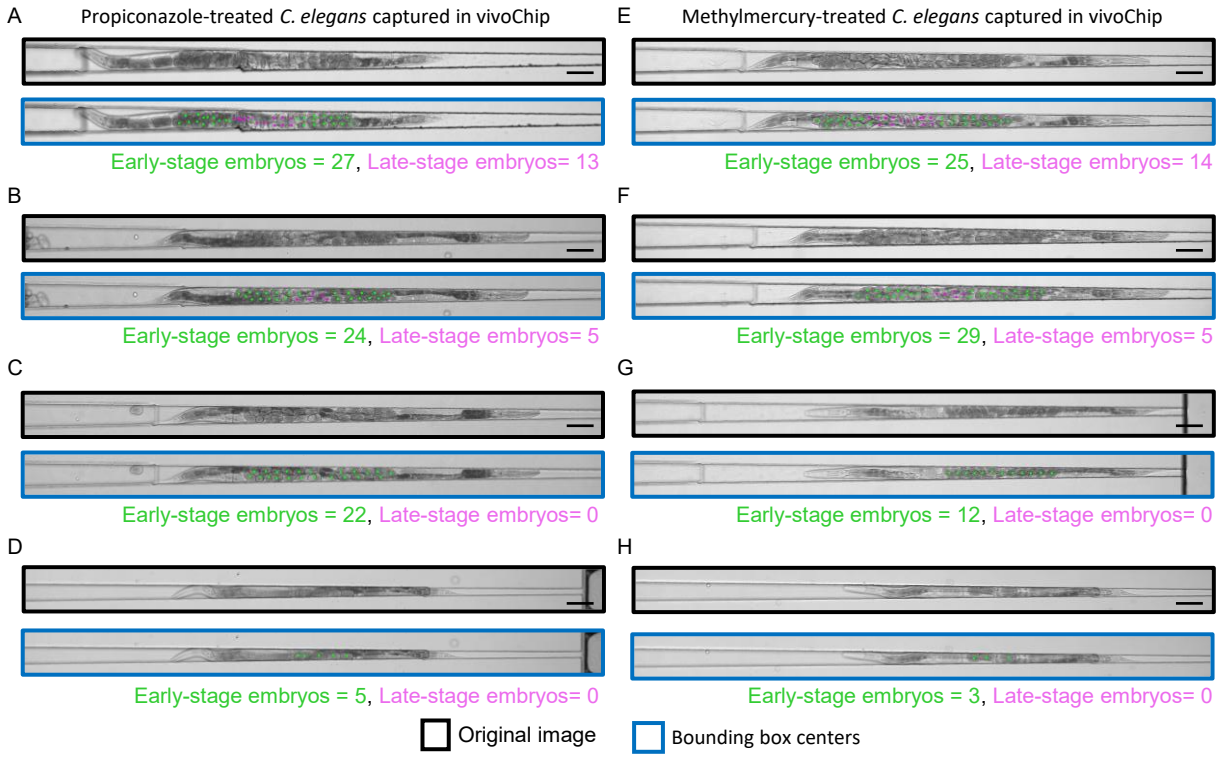
Supplementary Figure 4. Well-level average values for 3 embryo parameters from all 5 biological and 6 technical replicates. All 60 well-level average values from 30 wells (5 biological × 6 technical) are presented for each of the 4 DMSO concentrations. **(A-C)** Scatter plots of 30 well-level averages for the number of early-stage embryos (A), late-stage embryos (B), and total embryos (C) for four DMSO conditions. **(D-F)** Bar plot for well-level aggregated average values are plot for the number of early-stage embryos (D), late-stage embryos (E), and total embryos (F) for four DMSO conditions. Data obtained from EmbryoMAE-Det output (Model) and manually analyzed data (Manual) are presented. The red line indicates the mean ± SEM (n = 30) in A-C. The pair-wise comparison was performed, and the p-values are presented in **Supplementary Table 2**.



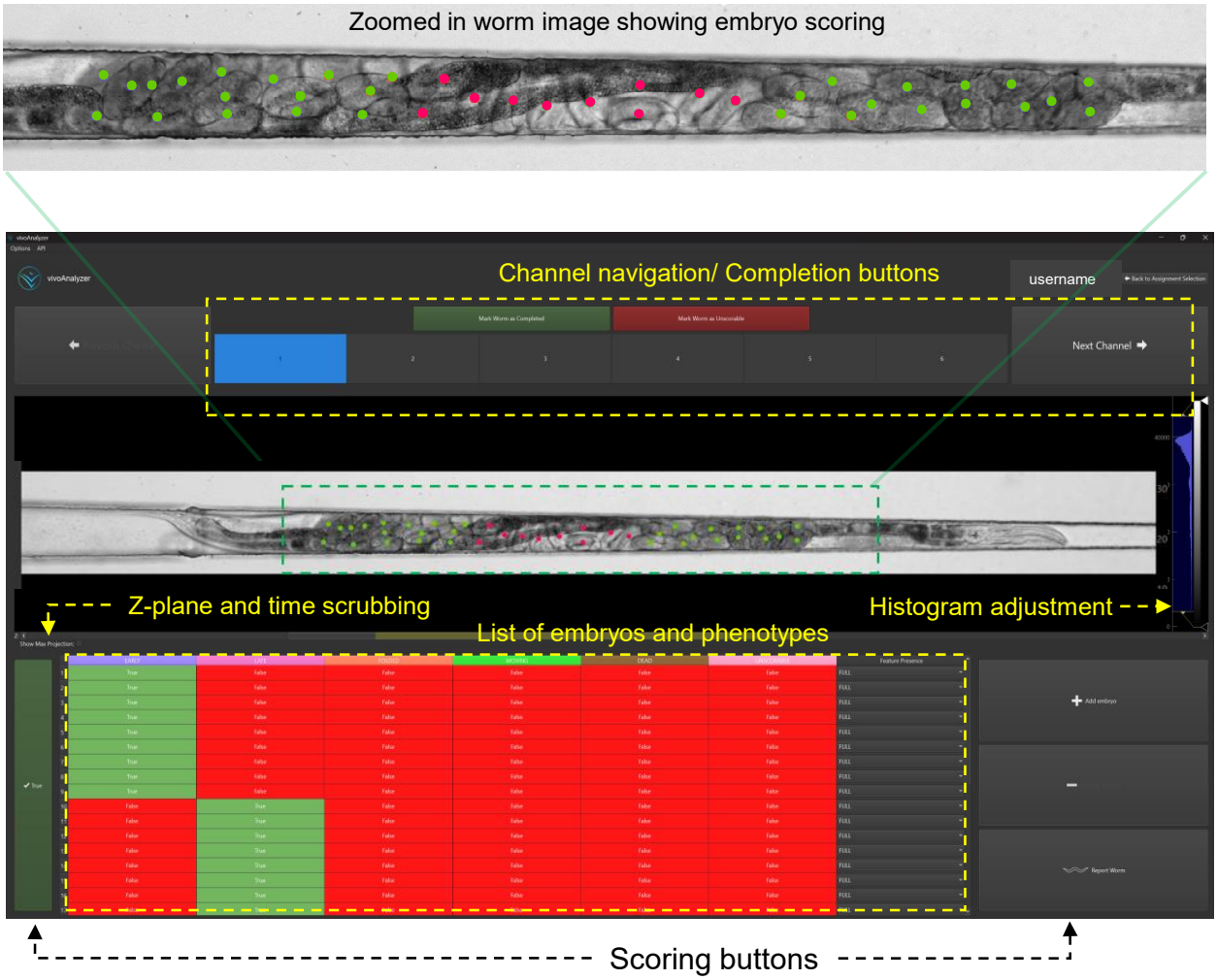
Supplementary Figure 5. Coefficient of variation from technical replicates. The coefficient of variation (CV%) for five biological replicates were calculated for 1% DMSO conditions and well averages from 6 technical replicates. The red lines represent mean ± SEM. The values obtained from model outputs were compared with manual analysis using paired t-test. The p-values ≥0.05 are denoted as not significant (ns).



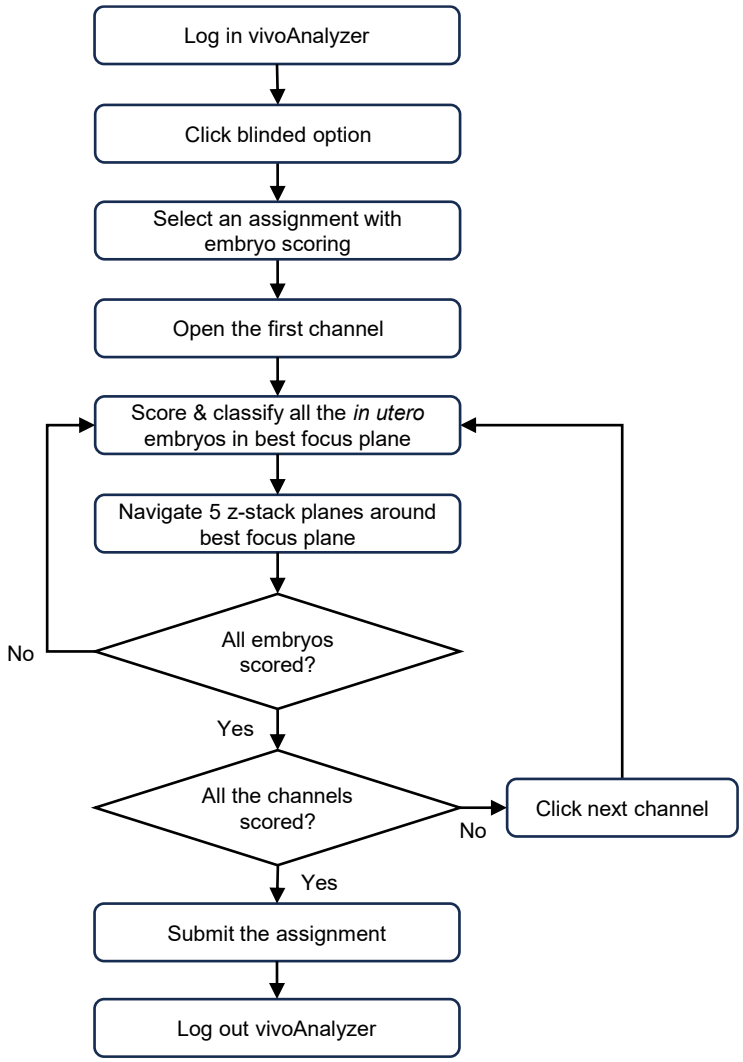
Supplementary Figure 6. Well-level average and CV% values for sixty 3-well combinations. For each DMSO condition, all 5 biological × 6 technical replicates were used to create sixty 3-well combinations. **(A-C)** Scatter plots of sixty 3-well combinations for early-stage embryos (A), late-stage embryos (B), and total embryos (C) across 4 DMSO conditions. **(D-F)** Scatter plots of sixty CV% for early-stage embryos (D), late-stage embryos (E), and total embryos (F) for 4 DMSO conditions. The red line indicates the mean ± SEM (*n* = 60). Values obtained from EmbryoMAE-Det output (Model) and with the manually analyzed data (Manual) are presented. The average values are presented in the main **Figs. 4C-H**. The pair-wise comparison was performed, and *p*-values are presented in **Supplementary Tables 3, 4**.



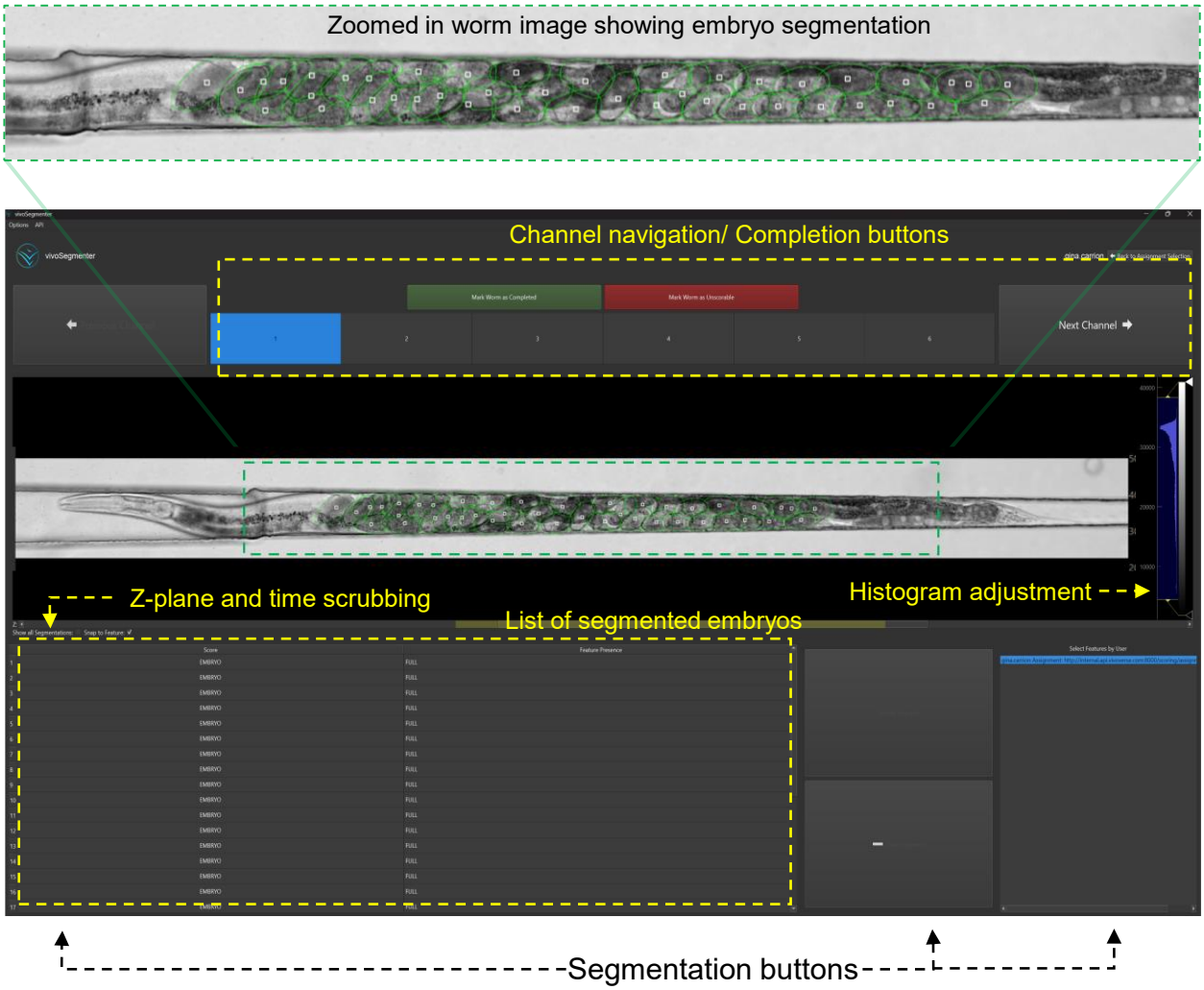
Supplementary Figure 7. Representative images of methylmercury-treated and propiconazole-treated worms. (A-D) Representative worm images treated with propiconazole in liquid media and imaged using vivoChip after 72 hours of exposure. **(E-H)** Representative worm images treated with methylmercury in liquid media and imaged using vivoChip after 72 hours of exposure. For each channel, the original brightfield image and the predicted embryo bounding box centers are shown. Based on the embryo box centers class, it is presented as early-stage embryos (green) and late-stage embryos (magenta). Scale bar is 200 µM.



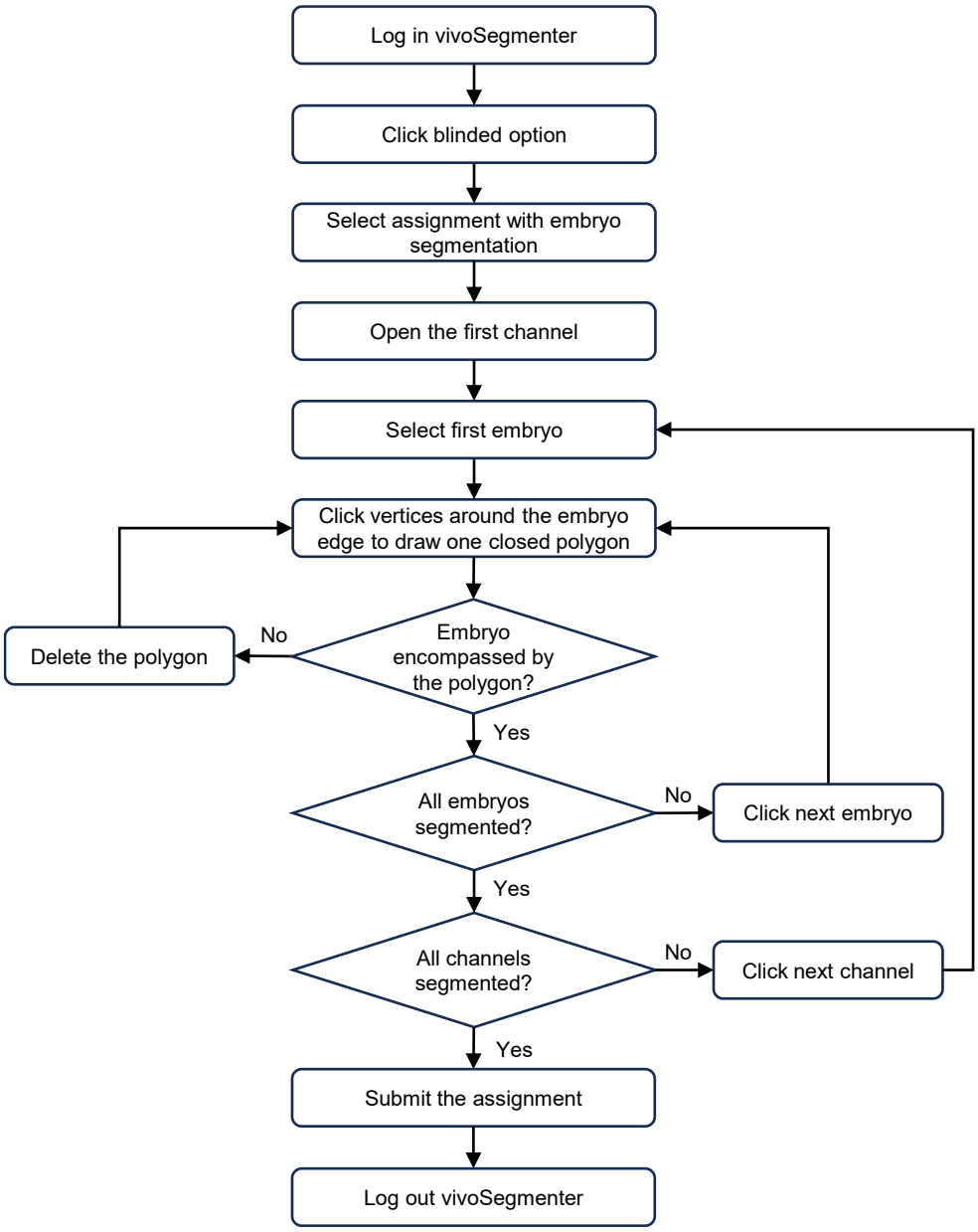
Supplementary Figure 8. Screenshot of vivoAnalyzer. Screenshot of vivoAnalyzer software. Users can navigate the experiment using the drop-down menu to select individual wells and microfluidic channels to visualize immobilized *C. elegans*. Intensity histograms can be adjusted, and multiple z-slices can be scanned to mark each embryo within the worm. The example image shows embryo scoring, with embryos classified into two classes: early (green dots) and late (pink dots). The top panel provides a zoomed-in view of the worm and its embryos.



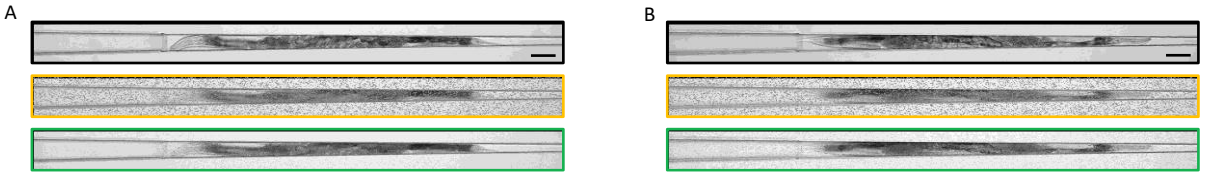
Supplementary Figure 9. Flow chart showing the steps involved in vivoAnalyzer software for embryo scoring. The software guides users through key steps, including loading images, scoring embryos, and saving scoring data for individual immobilized *C. elegans*. Scoring information stored in the database is subsequently used to train in-house machine learning models.



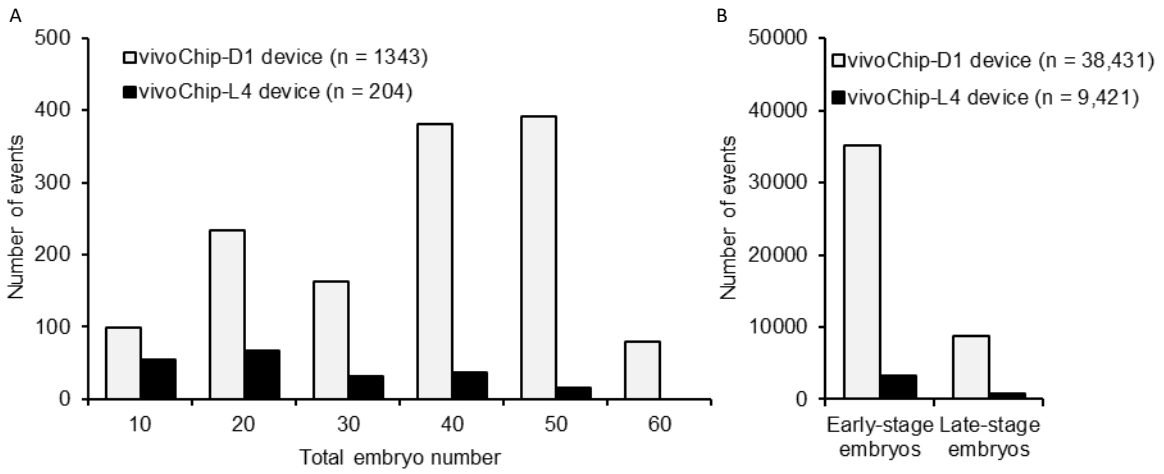
Supplementary Figure 10. Screenshot of vivoSegmenter software. Screenshot of vivoSegmenter software. Users can navigate the experiment using the drop-down menu to select individual wells and microfluidic channels to visualize immobilized *C. elegans*. Intensity histograms can be adjusted, and multiple z-slices can be scanned to define vertices for segmenting each embryo within a worm (green outlines). The main panel shows embryo segmentation of an immobilized *C. elegans*, while the top panel provides a zoomed-in view highlighting the segmented embryos.



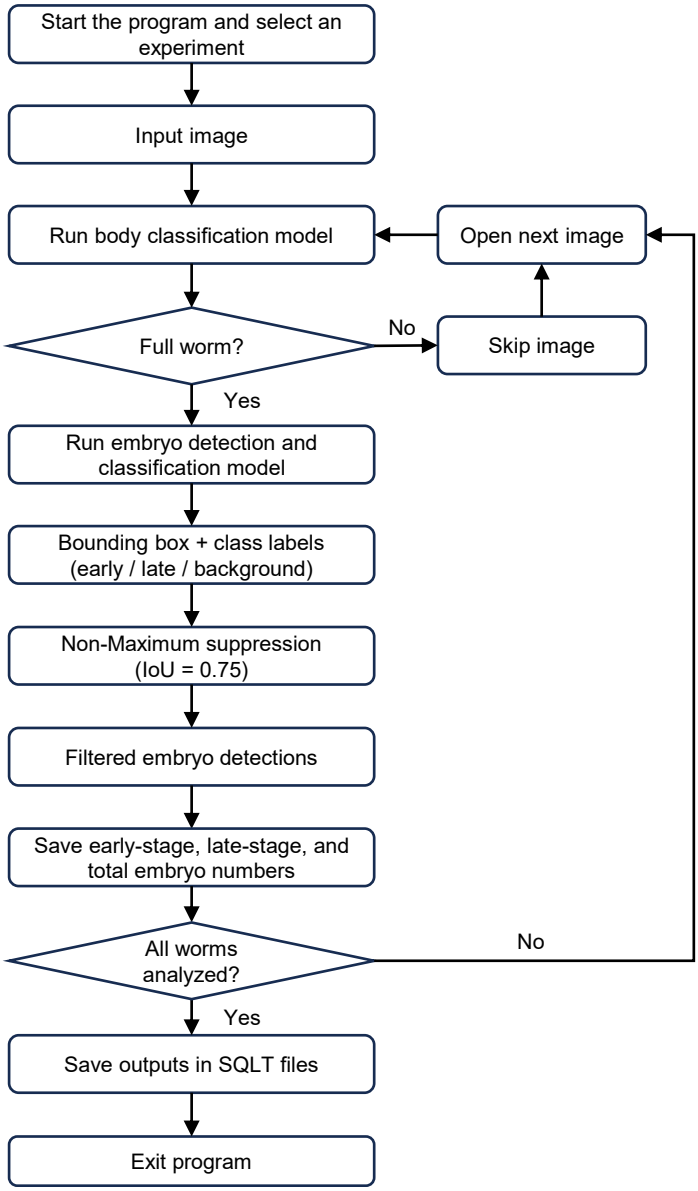
Supplementary Figure 11. Flow chart showing the steps involved in vivoSegmenter software for embryo segmentation. The software guides users through key steps, including loading images, segmenting embryos, and saving segmentation data for individual immobilized *C. elegans*. Segmentation information stored in the database is subsequently used to train in-house machine learning models.



Supplementary Figure 12. MAE reconstruction. **(A, B)** Two examples showing original images (black box), masked images with masked regions shown in gray (yellow box), and reconstructed images (output of pretrained model) from unmasked patches (green box). Scale bar is 200 µm.



Supplementary Figure 13. Training data used for embryo detection and classification model. The full training dataset consists of *n* = 1547 worms immobilized inside two types of vivoChip devices. **(A**) Histogram plot of worms with different minimum total numbers of embryos in their uterus imaged using vivoChip-24x-D1 (grey) and vivoChip-24x-L4 (black) devices. **(B)** Histogram plot of number of early-stage and late-stage embryos used for the training and collected with two types of vivoChip devices.



Supplementary Figure 14. Flowchart for end-to-end embryo inference pipeline. Each channel with *C. elegans* is analyzed for the embryo parameters following these steps. Only channels with full worms are used by the embryo model (EmbryoMAE-Det) to predict all the embryos inside the worm uterus. After all the channels are analyzed, the aggregated embryo parameters are stored in the vivoVerse database and as a SQLT file for further statistical analysis.

Supplementary Table 1. Model performance for embryo model ablation experiments.

Experiment	Avg precision (AP) for early-stage embryos	Avg precision (AP) for late-stage embryos	Mean average precision (mAP)
Experiment 1	79.5%	62.3%	70.9%
Experiment 2	82.4%	65.8%	74.1%
Final model (EmbryoMAE-Det)	92.8%	84.7%	88.7%

Supplementary Table 2. Statistical tests for well-level averages for 4 DMSO conditions. The aggregated well-level average values from all 30 wells (5 biological × 6-technical replicates) were calculated from both manual scores and model output values and compared.

Well-level averages	**0% DMSO (*p*-value)**	**0.2% DMSO (*p*-value)**	**0.5% DMSO (*p*-value)**	**1.0% DMSO (*p*-value)**
Early-stage embryos	<0.0001 (****)	<0.0001 (****)	<0.0001 (****)	0.0229 (*)
Late-stage embryos	0.1494 (ns)	0.0577 (ns)	0.0200 (*)	0.0064 (**)
Total embryos	<0.0001 (****)	<0.0001 (****)	<0.0001 (****)	<0.0001 (****)
Data was tested for normality using Kolmogorov-Smirnov test. Since several population did not follow normal distribution, we used Wilcoxon matched pairs signed rank test.				

Supplementary Table 3. Statistical tests for aggregated averages from 3-well combinations. The aggregated average from 3-well, 60 combinations were calculated from both manual scores and model output values and compared.

3-well, aggregated averages	**0% DMSO (*p*-value)**	**0.2% DMSO (*p*-value)**	**0.5% DMSO (*p*-value)**	**1.0% DMSO (*p*-value)**
Early-stage embryos	<0.0001 (****)	<0.0001 (****)	<0.0001 (****)	<0.0001 (****)
Late-stage embryos	0.0179 (*)	<0.0001 (****)	<0.0001 (****)	<0.0001 (****)
Total embryos	<0.0001 (****)	<0.0001 (****)	<0.0001 (****)	<0.0001 (****)
Tested for normality using Kolmogorov-Smirnov test. Since several population did not follow normal distribution, we used Wilcoxon matched pairs signed rank test.				

Supplementary Table 4. Statistical tests for aggregated CV% from 3-well combinations. The aggregated CV% values from 3-well, 60 combinations were calculated from both manual scores and model output values and compared.

Average CV%	**0% DMSO (*p*-value)**	**0.2% DMSO (*p*-value)**	**0.5% DMSO (*p*-value)**	**1.0% DMSO (*p*-value)**
Early-stage embryos	<0.0001 (****)	<0.0001 (****)	0.0017 (**)	0.0846 (ns)
Late-stage embryos	<0.0001 (****)	<0.0001 (****)	0.0008 (***)	0.0006 (***)
Total embryos	<0.0001 (****)	<0.0001 (****)	<0.0001 (****)	<0.0001 (****)
Tested for normality using Kolmogorov-Smirnov test. Since several population did not follow normal distribution, we used Wilcoxon matched pairs signed rank test.				

Supplementary Table 5. Effective concentration values for embryo-related parameters for propiconazole-treated worms. The effective concentration (EC_{50}) and 95% confidence intervals (95% CI) values for 3 embryo parameters are presented. The concentration-response curves for model output and manual data were compared for similar EC_{50} values and Hillslope.

Parameter	EC_{50} (95% CI) from model output	EC_{50} (95% CI) from manual scoring	F-value (DF1, DF2), *p*-value
Early-stage embryo	265 µM (242-292)	255 µM (234-279)	0.2189 (2, 54), 0.8041 (ns)
Late-stage embryo	84 µM (36-160)	79 µM (35-154)	0.0196 (2, 54), 0.9806 (ns)
Total embryo	214 µM (187-246)	207 µM (179-239)	0.0740 (2, 54), 0.9288 (ns)

Supplementary Table 6. Effective concentration values for embryo-related parameters for methylmercury-treated worms. The effective concentration (EC_{50}) and 95% confidence intervals (95% CI) values for 3 embryo parameters are presented. The concentration-response curves for model output and manual data were compared for similar EC_{50} values and Hillslope.

Parameter	**EC_{50} (95% CI) from model output**	**EC_{50} (95% CI) from manual scoring**	**F-value (DF1, DF2), *p*-value**
Early-stage embryo	3.90 µM (3.66-4.15)	3.90 µM (3.66-4.16)	0.0077 (2, 110), 0.9923 (ns)
Late-stage embryo	1.41 µM (1.28-1.56)	1.49 µM (1.33-1.67)	0.2589 (2, 110), 0.7723 (ns)
Total embryo	3.16 µM (2.90-3.43)	3.16 µM (2.91-3.42)	0.0465 (2, 110), 0.9546 (ns)