

Representation Learning of Zebrafish Embryo Morphodynamics

Autoencoders + Latent Space Analysis (PCA & UMAP)

Pablo Escobar

Yachay Tech

December 16, 2025

Motivation

- ▶ Early development is dynamic: morphology changes continuously over time.
- ▶ Manual phenotyping is slow and sensitive to subjective bias.
- ▶ Goal: learn a compact, quantitative *morphospace* directly from images (unsupervised).

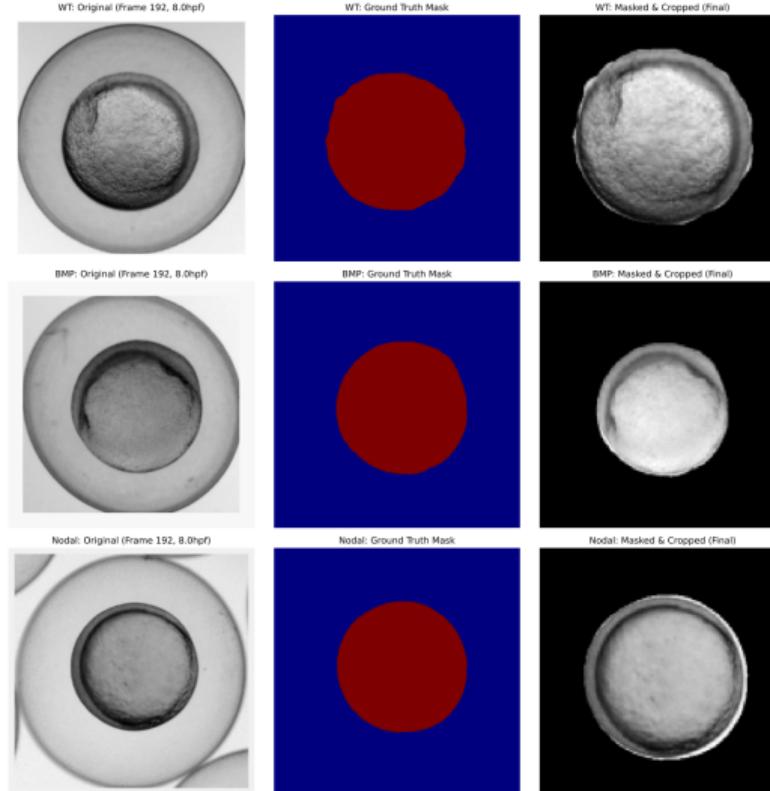
Introduction & Research Question

- ▶ Dataset: zebrafish embryos, 2–16 hpf, WT vs BMP-perturbed vs Nodal-perturbed.
- ▶ Question: can representation learning capture developmental trajectories and detect mutant divergence early?
- ▶ Strategy: preprocessing → autoencoder learning → reconstruction check → PCA/UMAP latent analysis.

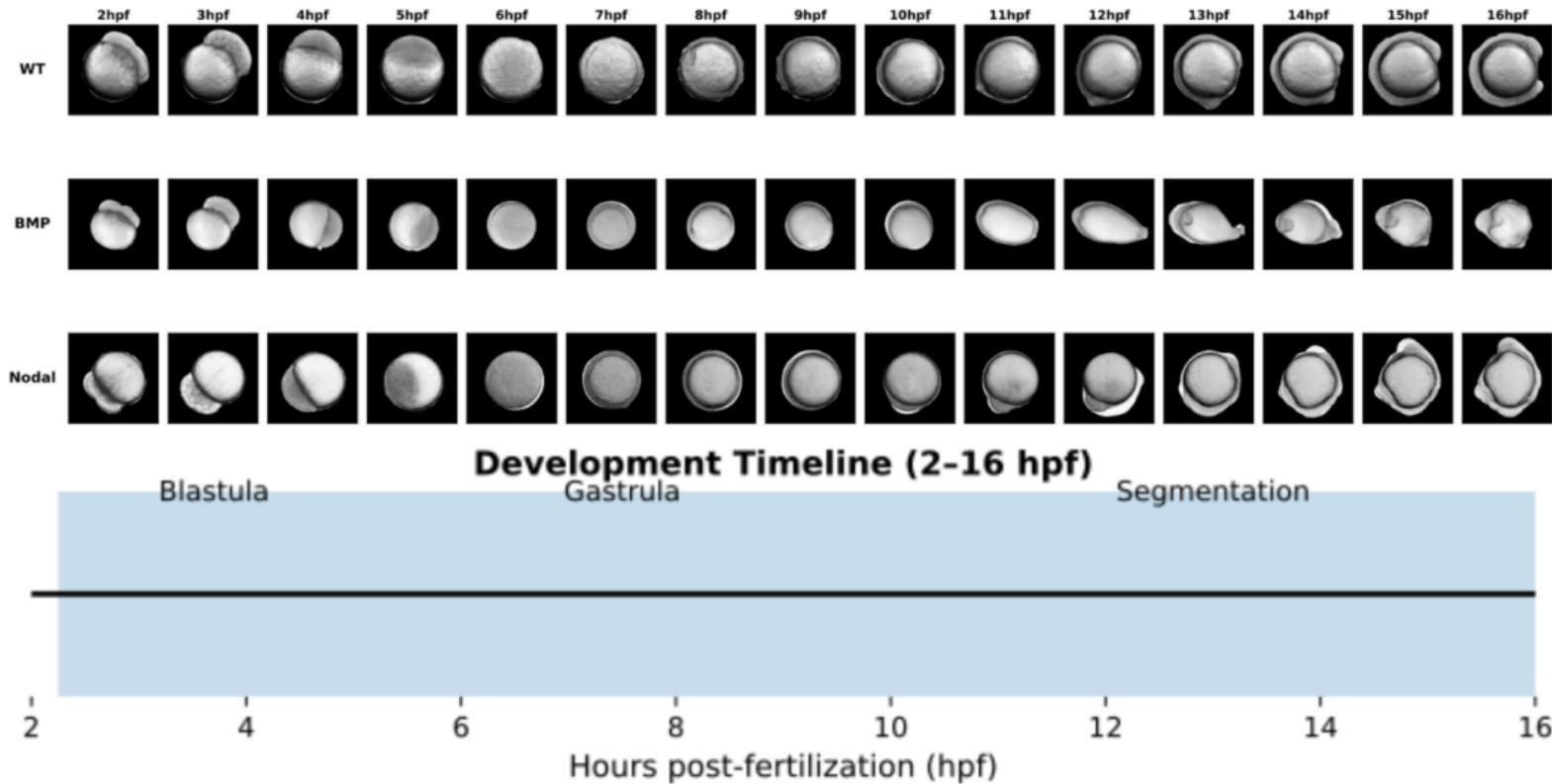
Contributions

- ▶ Standardized preprocessing pipeline for temporally aligned embryo stacks.
- ▶ Unsupervised autoencoder representation (128-D latent code) for morphology encoding.
- ▶ Quantitative evaluation: reconstruction error across time and genotype.
- ▶ Latent-space results: genotype clustering, trajectories, and divergence timing.

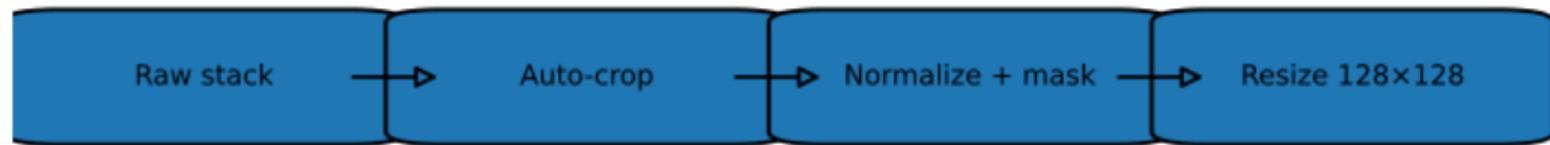
Dataset & Preprocessing Overview: Raw → Mask → Final



Dataset & Preprocessing Overview: Development Timeline (2–16 hpf)

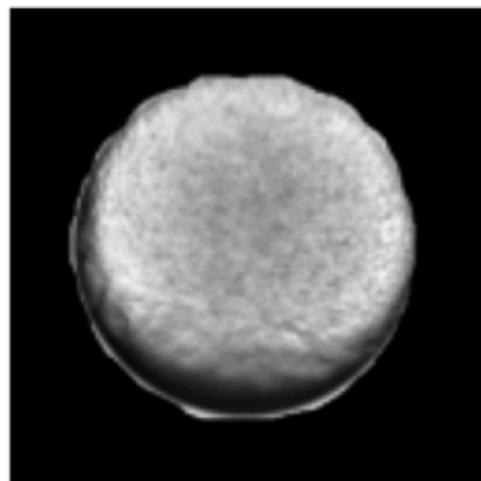


Dataset & Preprocessing Overview: Preprocessing Pipeline

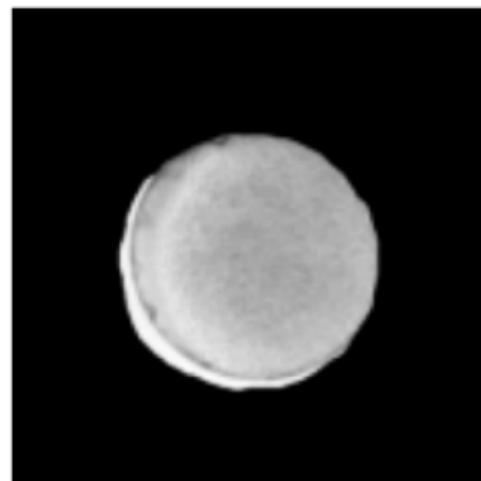


Dataset & Preprocessing Overview: Representative Development (WT/BMP/Nodal)

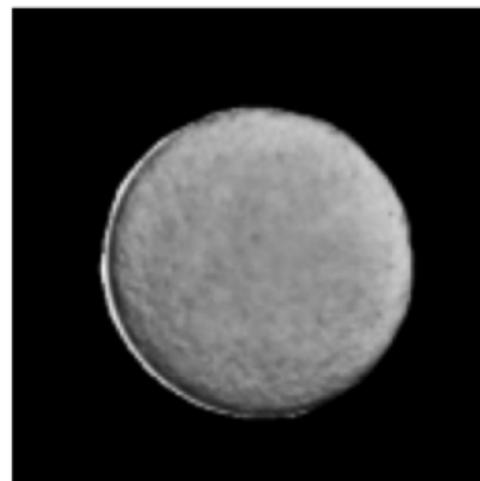
WT



BMP



Nodal



Autoencoder Architectures: Data Splits (Train / Val / Test)

Training Data Batch (Augmented)



Validation Data Batch (Base Transform)

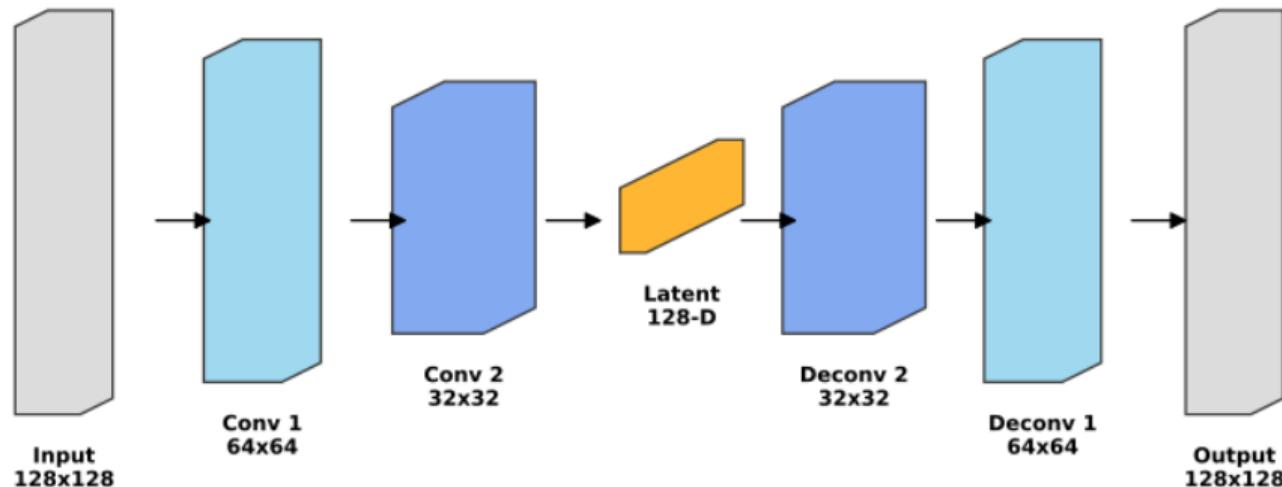


Testing Data Batch (Base Transform)

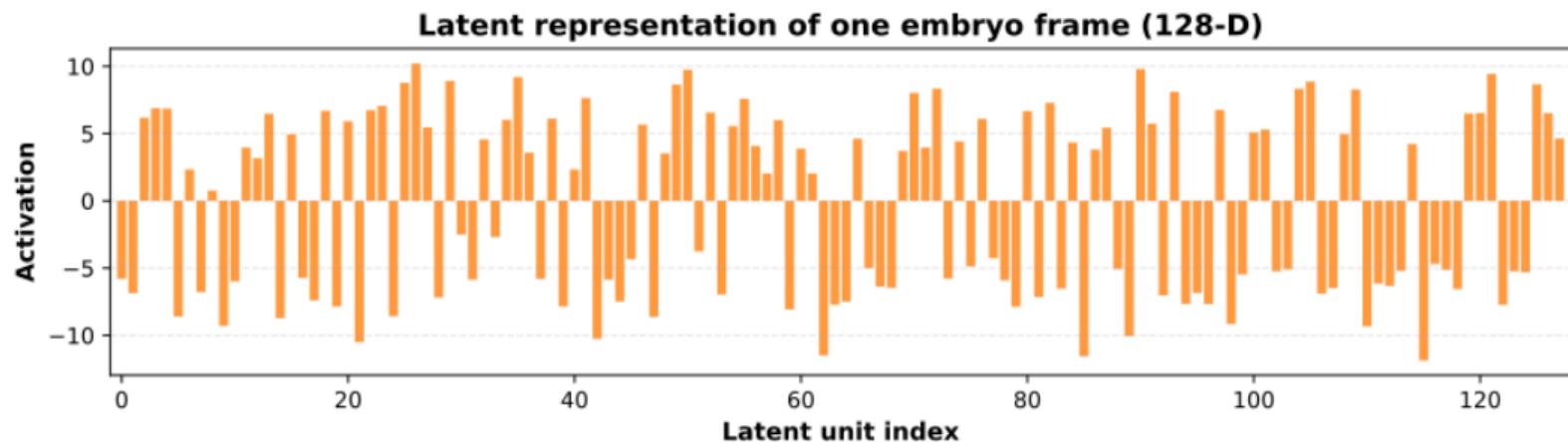


Autoencoder Architectures: Unsupervised AE (128-D Latent)

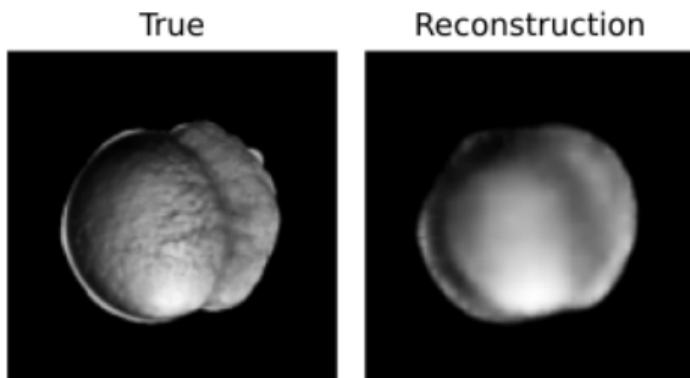
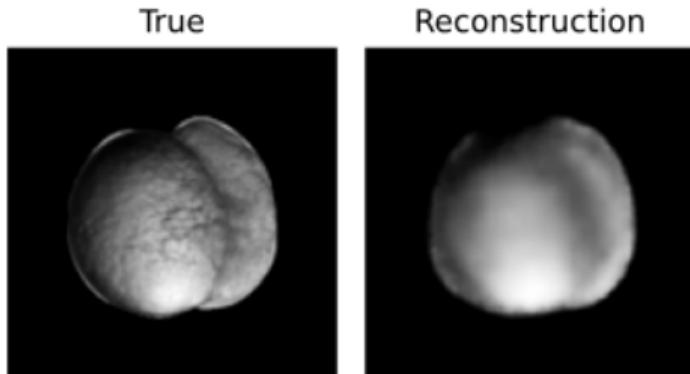
Zebrafish Autoencoder Architecture (Unsupervised)



Latent Dimension Illustration (128-D)



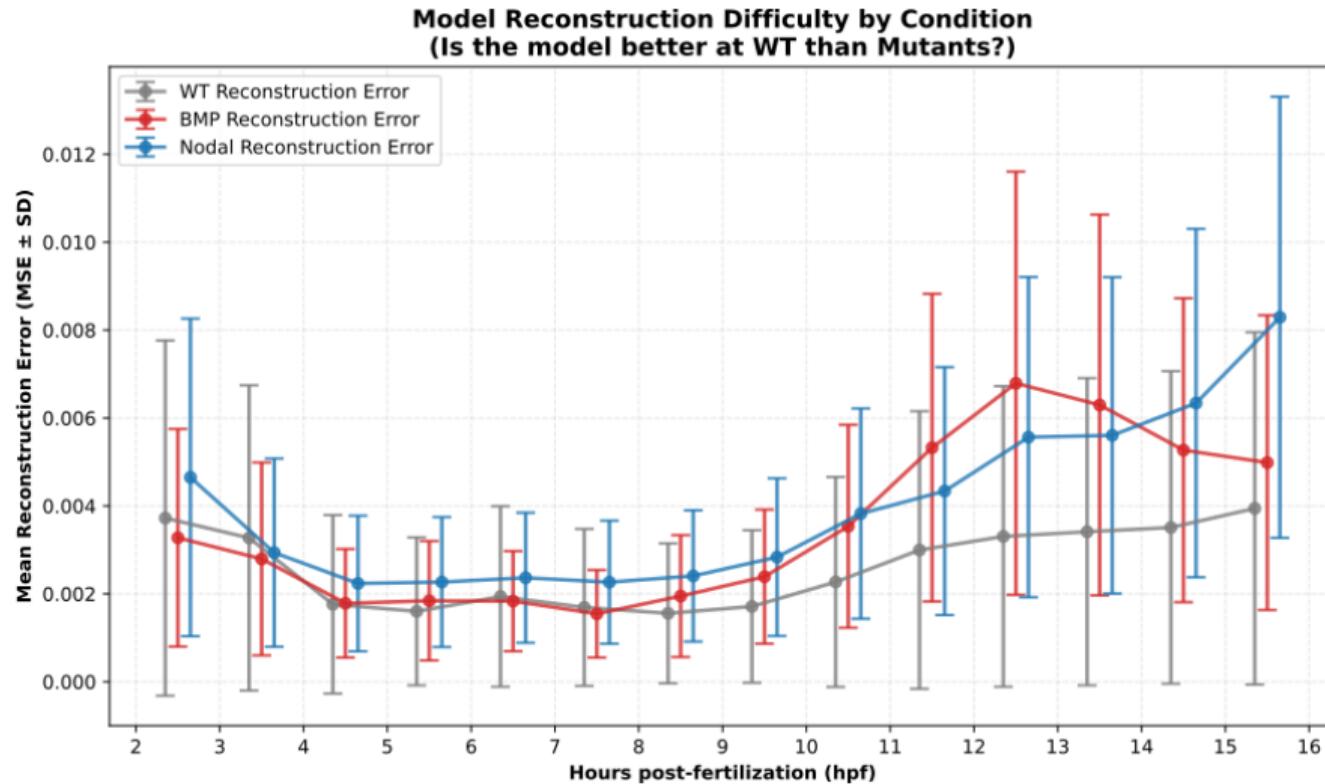
Reconstruction Results: True vs Reconstructed



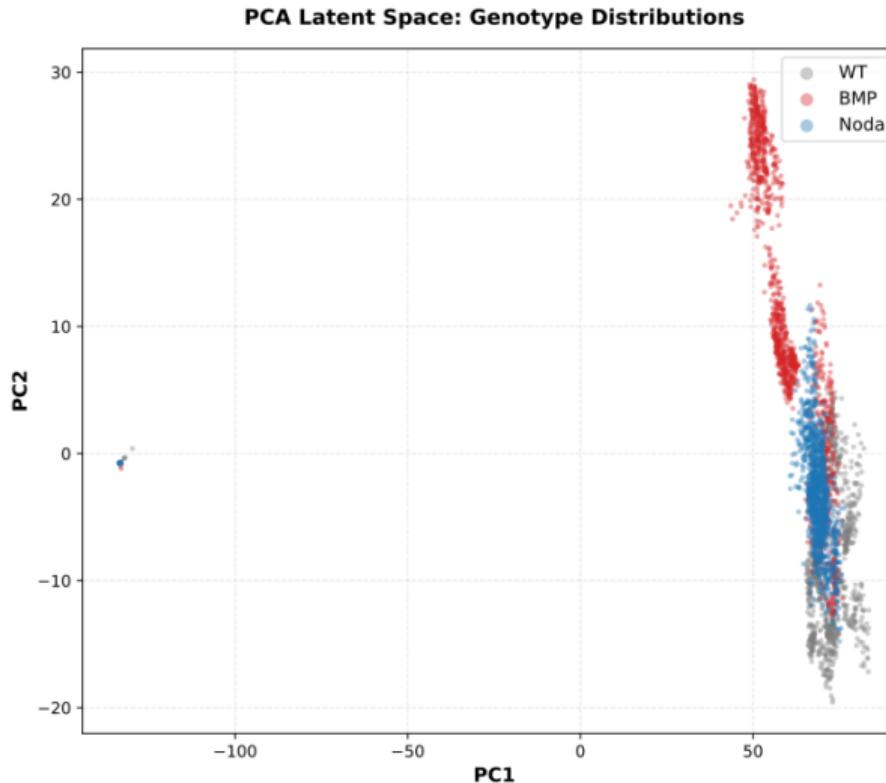
Reconstruction Error Over Development (2–16 hpf)



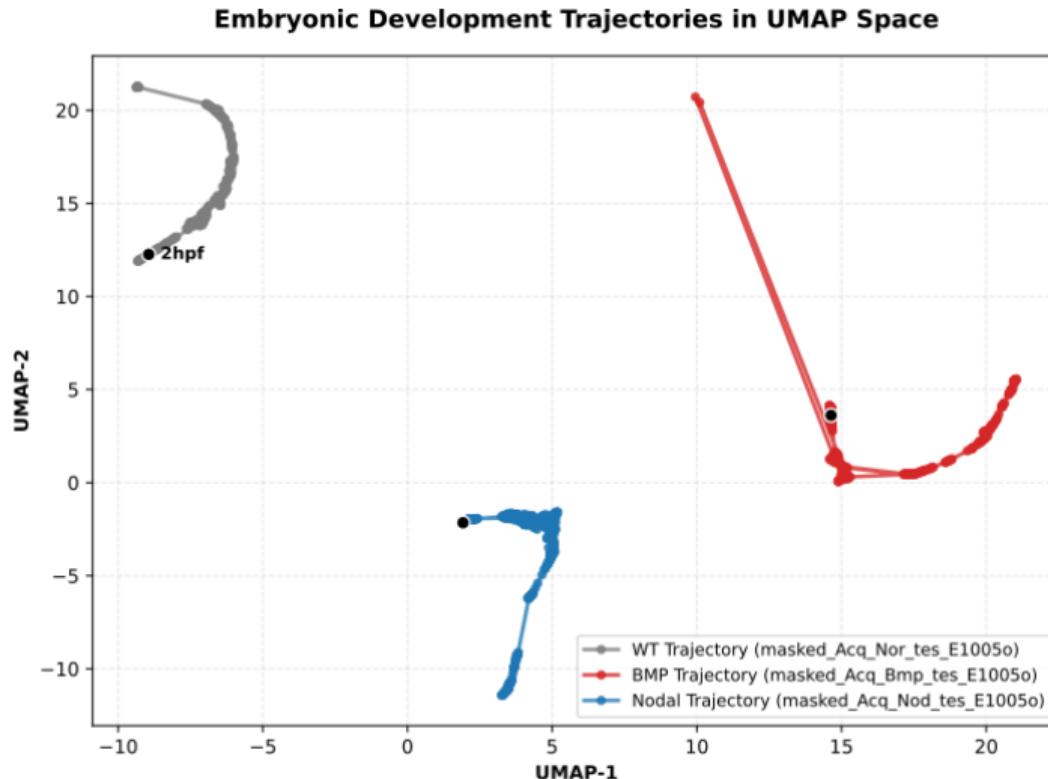
Reconstruction Difficulty by Condition (WT vs BMP vs Nodal)



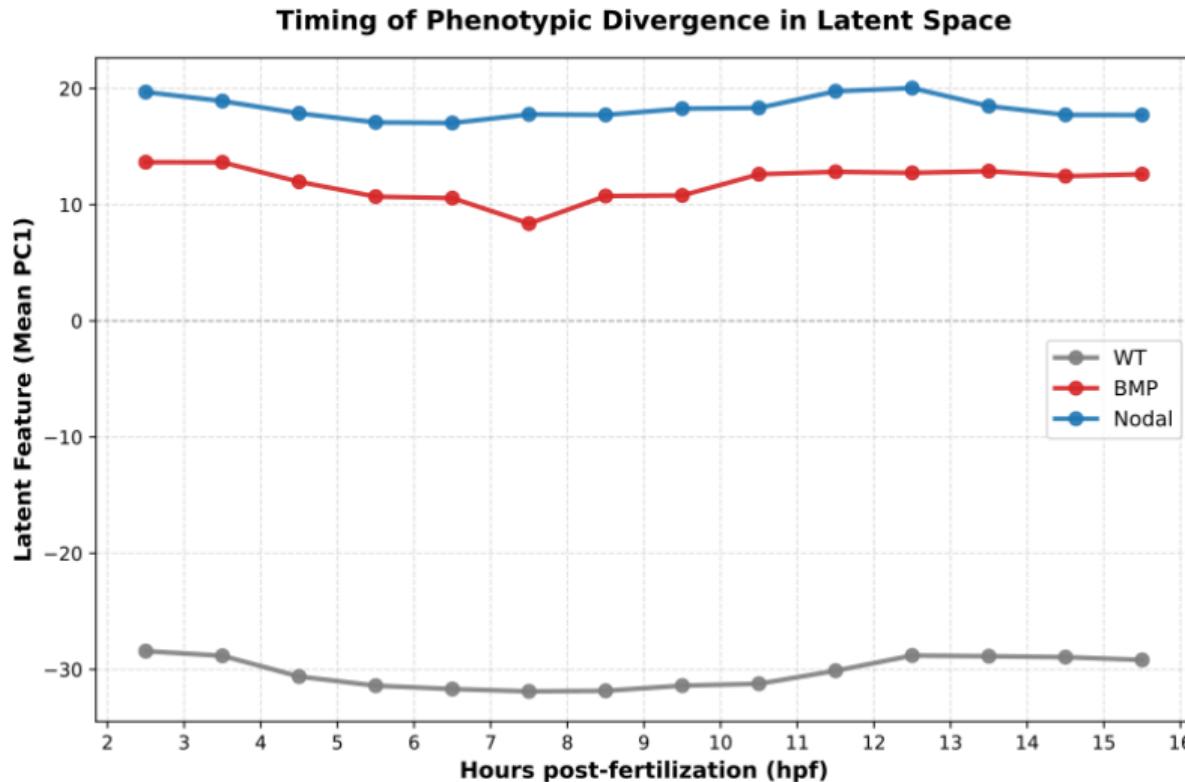
Latent Space Organization: PCA Embedding by Genotype



Latent Space Organization: UMAP Trajectories Over Time

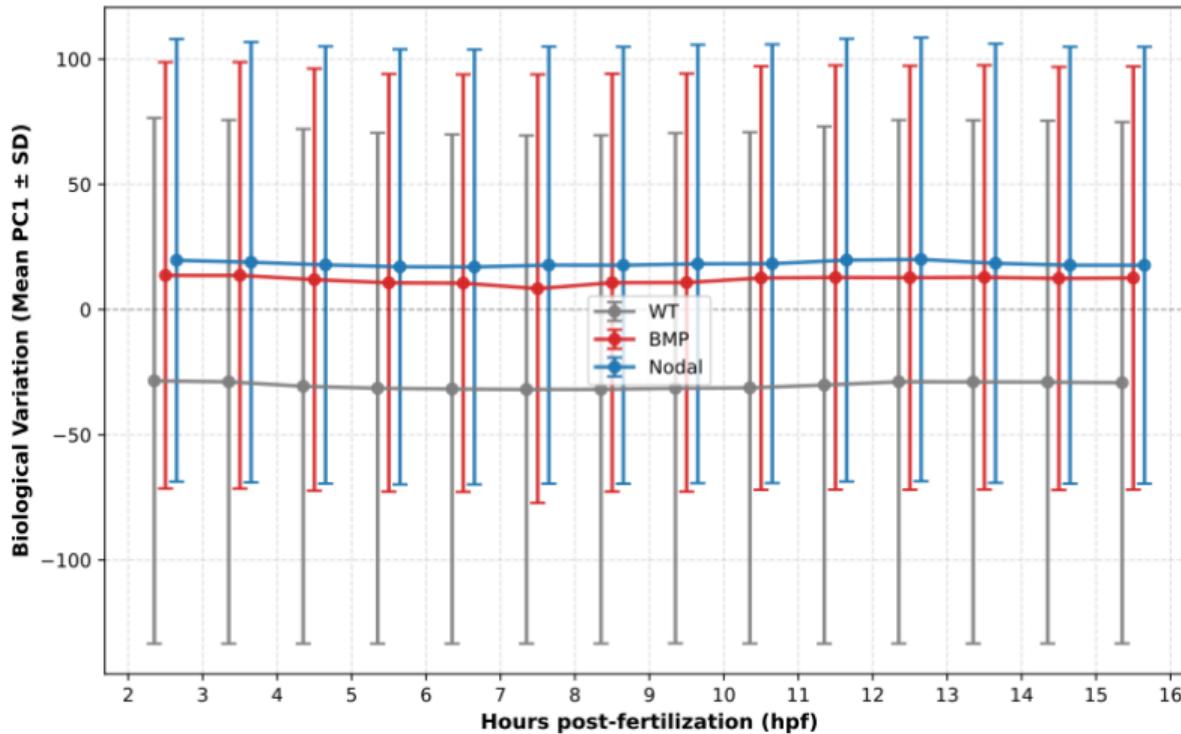


Timing of Phenotypic Divergence in Latent Space (PC1 vs hpf)



Timing of Phenotypic Divergence (Mean \pm SD)

Timing of Phenotypic Divergence in Latent Space



Summary of Findings / Conclusions

- ▶ Preprocessing yields consistent, background-free inputs across 2–16 hpf.
- ▶ Autoencoder learns a compact 128-D morphospace with preserved structure.
- ▶ Reconstruction error increases with developmental complexity and is higher for mutants.
- ▶ Latent space separates WT/BMP/Nodal and reveals genotype-specific divergence timing.

Thank you

Questions?