

Metrics for Comparing Cell Lineages

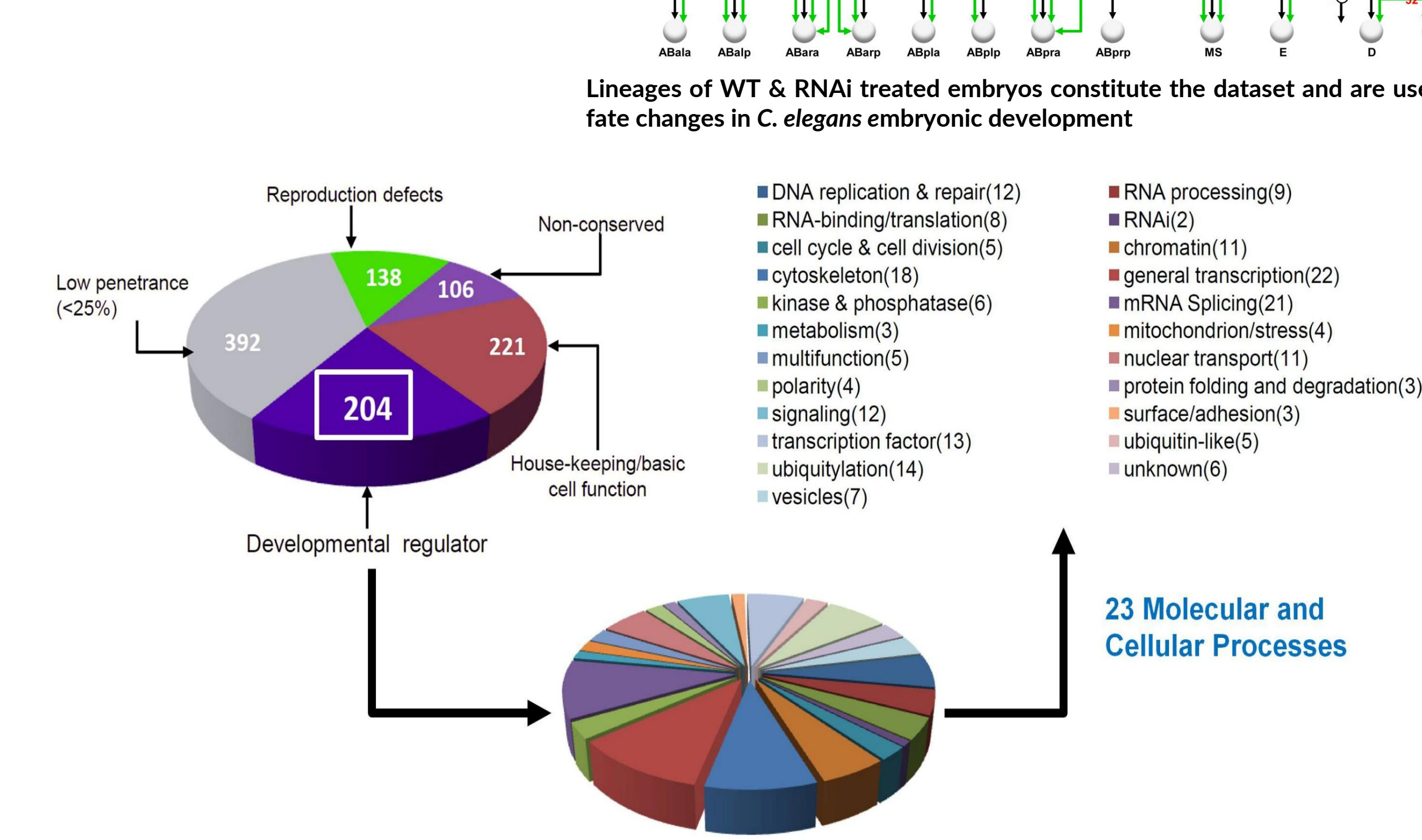
Gunalan Natesan^{1,3}, Pavak K. Shah^{1,2}, Eric J. Deeds^{1,3}

- 1. Computational and Systems Biology IDP, University of California, Los Angeles (UCLA), Los Angeles, CA, USA
- 2. Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, (UCLA), Los Angeles, CA, USA
- 3. . Department of Integrative Physiology and Bioinformatics, University of California, Los Angeles, (UCLA), Los Angeles, CA, USA

BACKGROUND: All multicellular organisms develop through the proliferation and differentiation of a single progenitor cell. Here we propose a pair of metrics derived from graph theory to define distances between cell lineages recorded from developing embryos by live microscopy. Using a previously published resource of cell lineages recorded from wild-type and RNA-perturbed embryos of the nematode *C. elegans*, which possesses an invariant cell lineage, we demonstrate that our metrics can reveal previously unappreciated batch effects in lineage recordings and offer new insights into the structure of cell lineages as a phenotypic space.

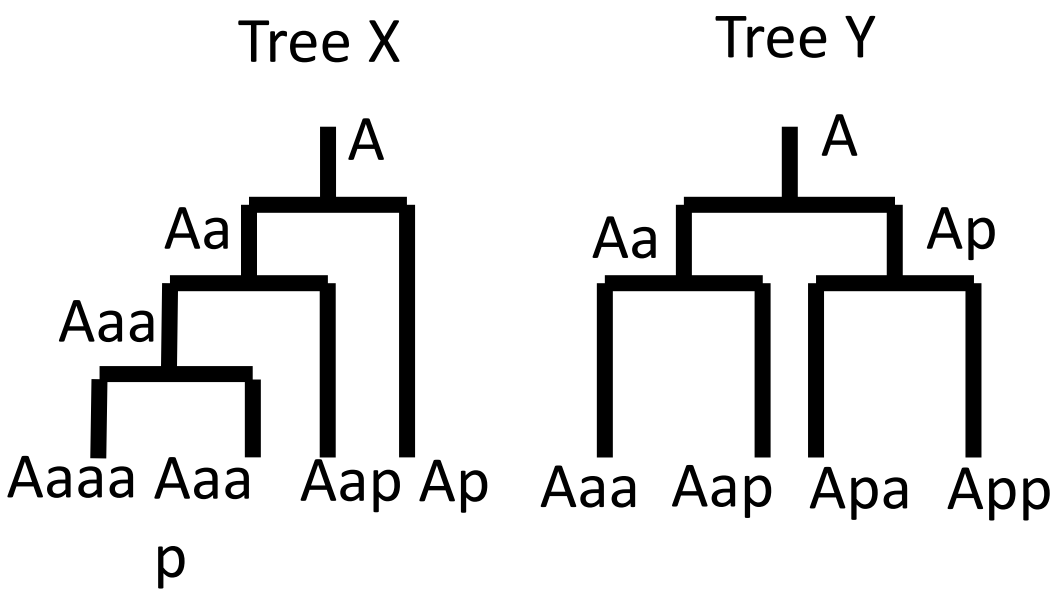
OUR DATASET

A comprehensive RNAi screen of 1061 *C. elegans* genes required for viability found 204 developmental regulators regulating 23 processes with broad cellular and molecular functions



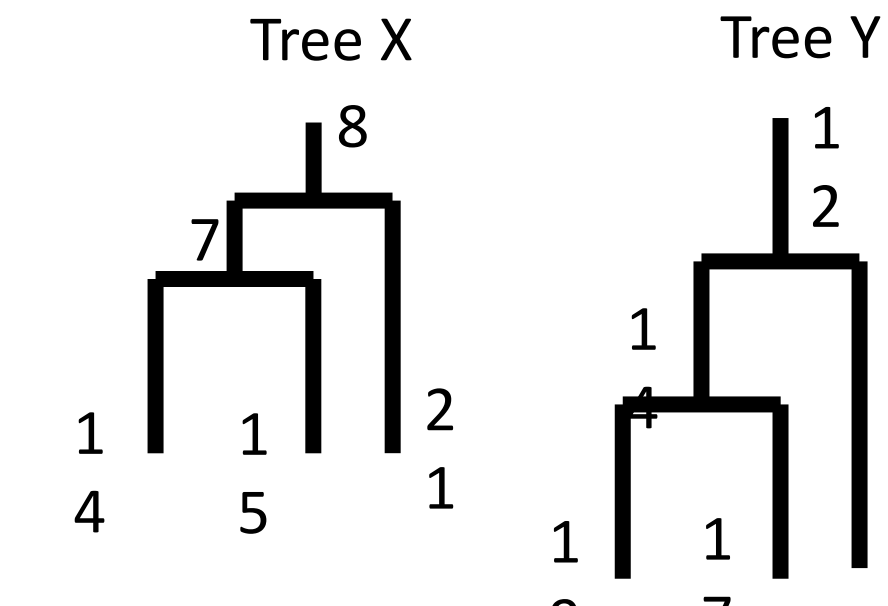
PROPOSED METRICS

We thus propose the usage of 2 unique graph metrics:



Tree Edit Distance represents differences in the number and placement of nodes between Trees. TED between Trees X and Y finds the size of set XOR(X,Y)

$$TED = \left| \begin{matrix} X \\ A, Aa, Aaa, \\ Aaaa, Aaap, \\ Aap, Ap \end{matrix} \right| \text{ XOR } \left| \begin{matrix} Y \\ Aa, Aaa, Aap, \\ A Ap, Apa, \\ App, \end{matrix} \right| = \left| \begin{matrix} XOR(X,Y) \\ Aaaa, \\ Apa, App \\ Aaap \end{matrix} \right| = 4$$

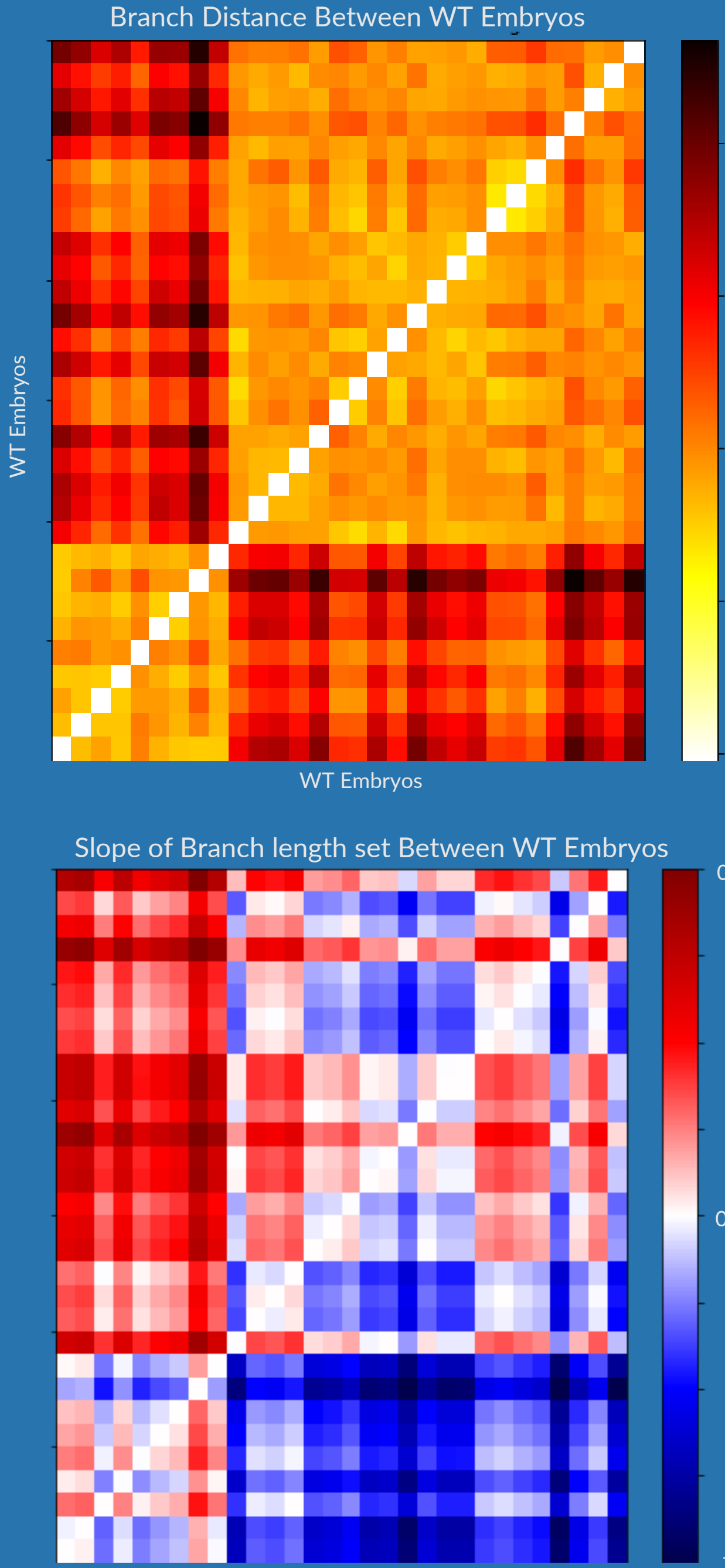


Branch Distance represents differences between edge lengths of Trees. BD between Tree X and Y finds the sum of squared differences between corresponding edges

$$BD = \sqrt{(8-12)^2 + (7-14)^2 + (14-19)^2 + (15-17)^2 + (21-31)^2} = 13.93$$

Our metrics reveal unrecognized heterogeneity in cell lineages during embryogenesis

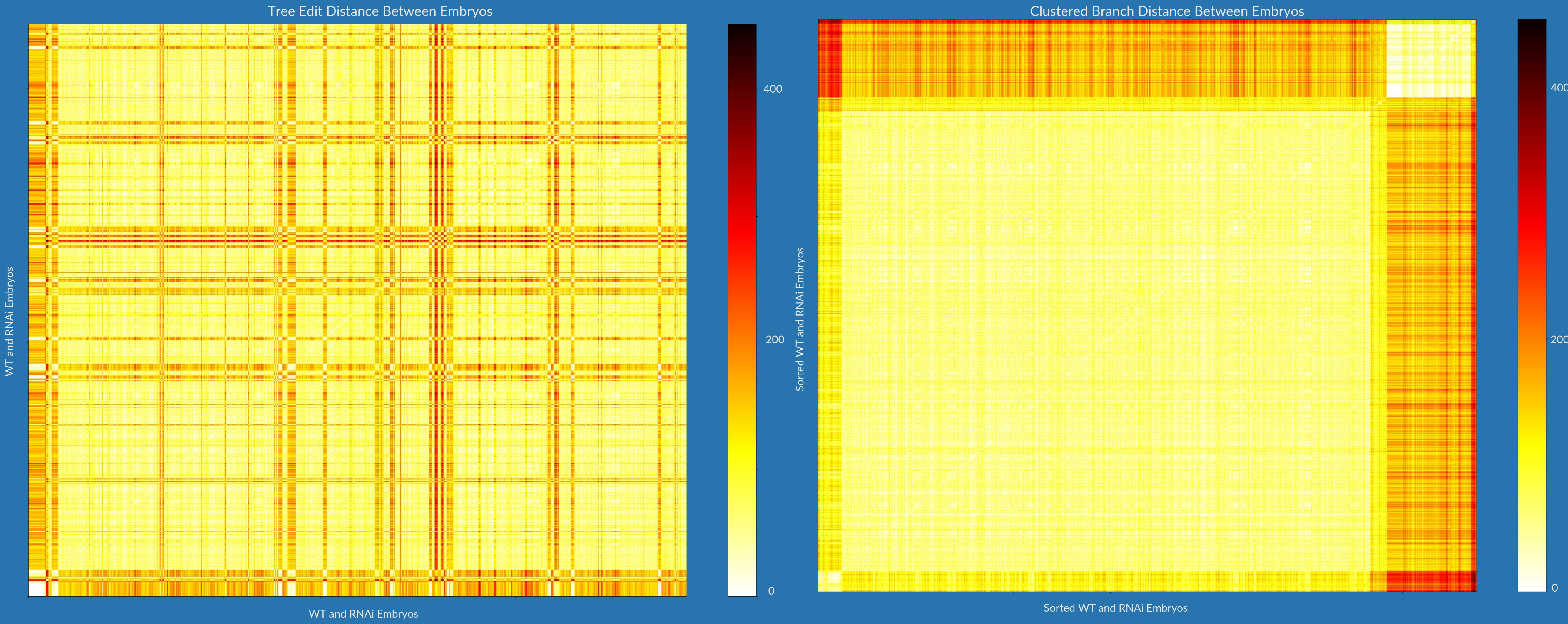
Branch edit distance detects batch effects in wild type embryogenesis due to global scaling



The Branch Distance between all WT embryos is computed, clustered, and analyzed to reveal two distinct phenotypes. These reveal more heterogeneity in the WT developmental clock than expected in this dataset.

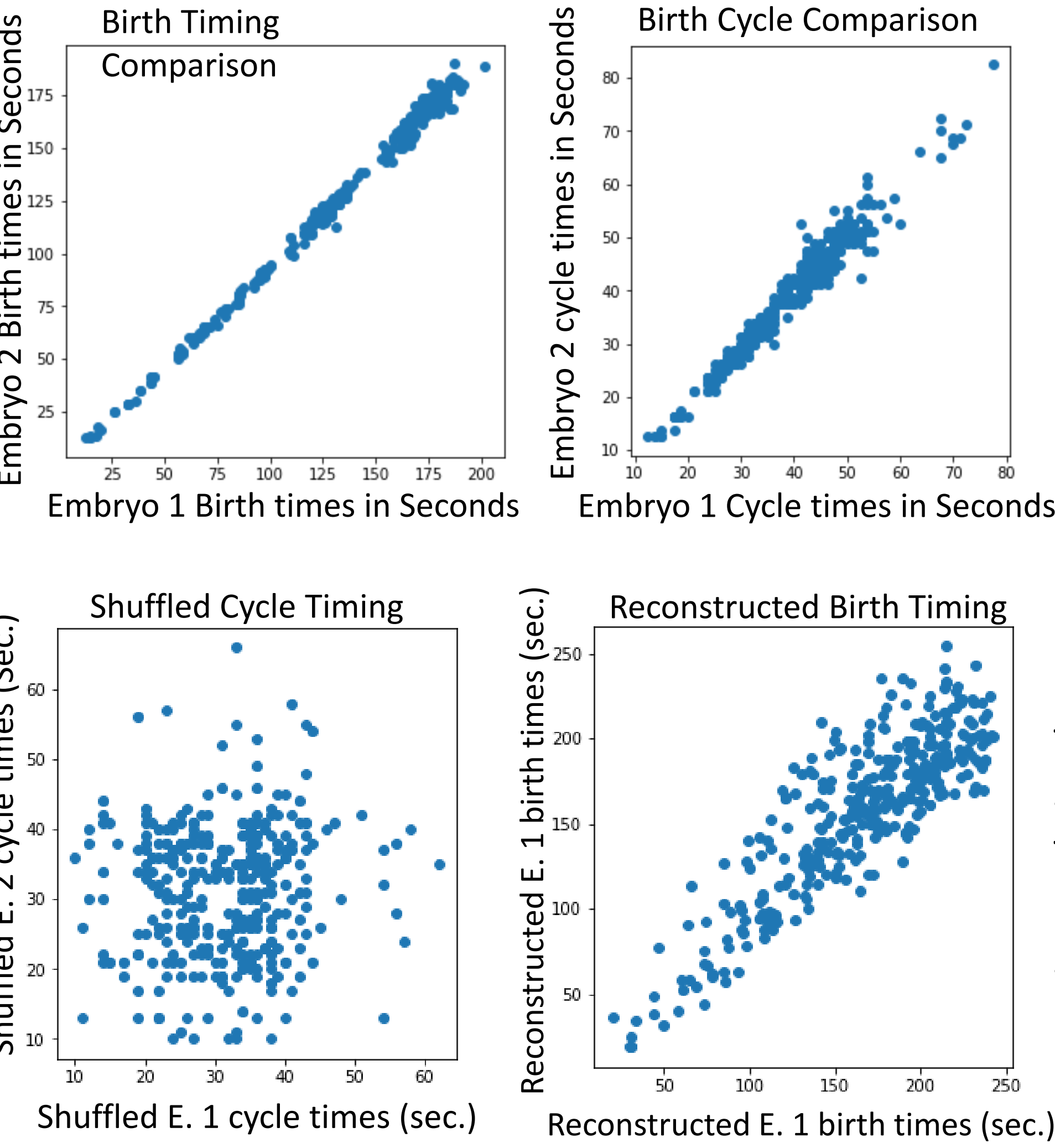
The slope between corresponding branch lengths in WT embryos are compared. The opposing sign between each batch suggest that each is a linearly scaled version of the other, indicating a difference in the global clock rate.

Tree-based distances define a global landscape of heterogeneity in developmental lineages under genetic perturbation



Each WT or Experimental embryo is a certain distance (TED/BD) from all other embryos. How do distances reflect known fate changes?

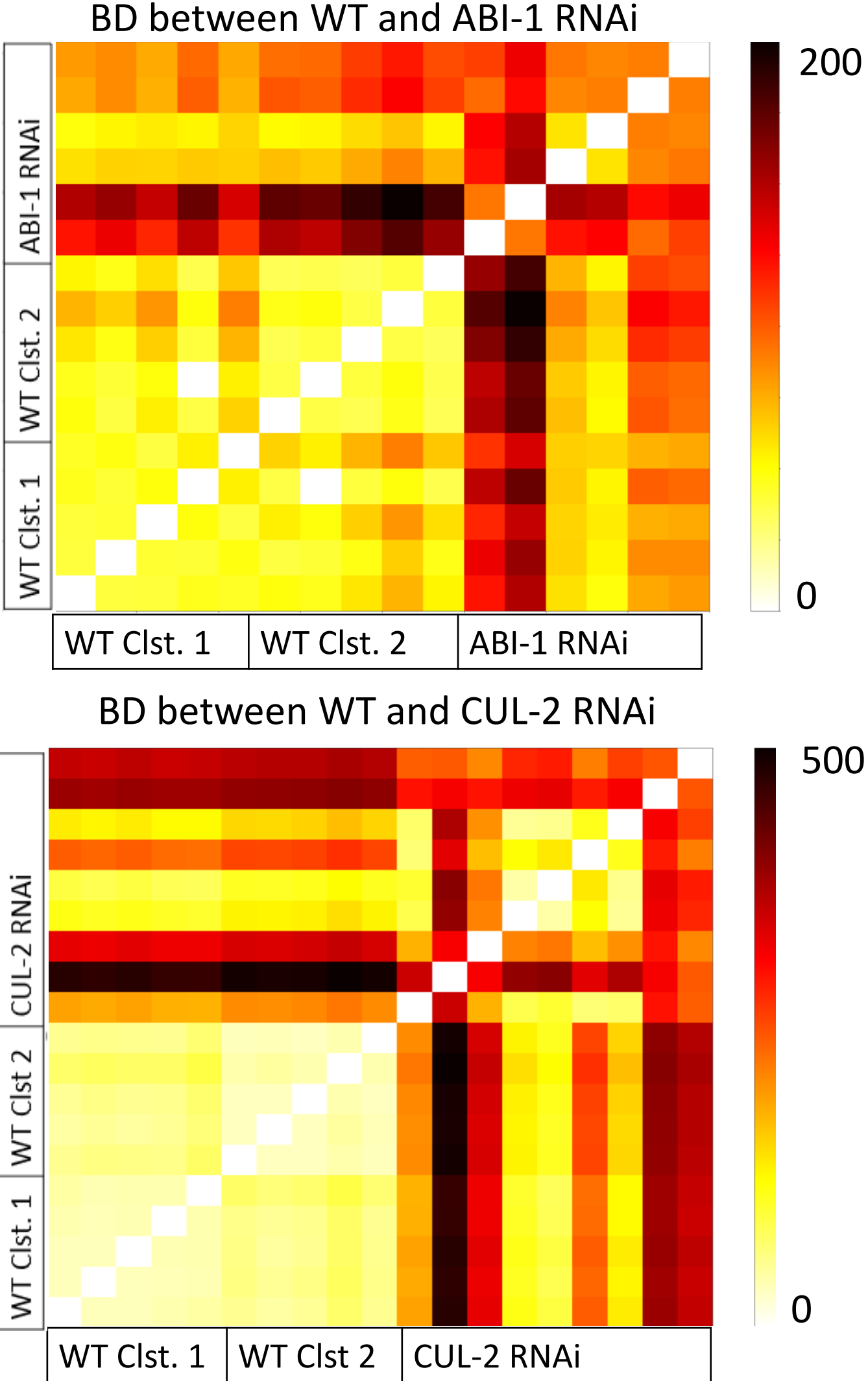
HOW TO MEASURE TIMING?



Birth times of each cell in an embryo are plotted [left] against another embryo to compare developmental clocks. Birth time is often used but typically appears less variable than the duration of each cell cycle[right]. Is there a reason to use on over the other?

Birth times are the sums of cycle times and are thus subject to the law of averages. This is shown [left] where the cell cycle times are shuffled ($r^2 = 0.001$) as a baseline for their summation [right] ($r^2 = .7492$). Change in r^2 is from mean regression under sums, masking heterogeneity

DEFINING THE PHENOTYPIC LANDSCAPE OF RNAi-INDUCED PERTURBATIONS



- RNAi can result in incomplete and variable knockdown of target gene transcripts
- Does variation in gene dose cause variation in the severity of developmental phenotypes?
- Our graph metrics suggest that rather than simply variability in the severity of phenotypes, many perturbations result in qualitatively different phenotypes

REFERENCES

Du, Zhuo, et al. "The Regulatory Landscape of Lineage Differentiation in a Metazoan Embryo." *Developmental Cell*, vol. 34, no. 5, 2015, pp. 592–607., <https://doi.org/10.1016/j.devcel.2015.07.014>.

Schwarz, Stefan et al. "A New Perspective on the Tree Edit Distance." *SISAP* (2017).

Johnson, S C. "Hierarchical clustering schemes." *Psychometrika* 32 (1967): 241-254.

Scikit-learn: *Machine Learning in Python*, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

Pauli Virtanen et al, and SciPy 1.0 Contributors. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17(3), 261-272.

Harris, C.R., Millman, K.J., van der Walt, S.J. et al. *Array programming with NumPy*. Nature 585, 357–362 (2020). DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2)