Fractal-Like Density in Single Cell Data

Timothy Hamilton[1,2,3], Breanne Sparta[2,3], and Eric J. Deeds[2,3]

***Short Abstract* — Waddington's Landscape is an influential paradigm describing how single cells develop and differentiate into their mature forms. For over 60 years, the landscape has been quantitatively characterized as representing a dynamical system built on and on top of gene expression, with cell types corresponding to regions around attractors in the dynamics of the underlying gene regulatory network. With the advent of scRNA-seq, it was thought that these attractor regions would be seen in the high-dimensional space of mRNA levels. Our analysis shows, however, that this isn't the case, recontextualizing expected heterogeneity amongst single cells during development**

***Keywords* — single cell transcriptomics, cell types, Waddington's Landscape, Scale-free networks, graph theory**

I. Introduction

Wadddington's Landscape has played an integral role in the interpretation of the developmental process that organisms undergo to develop specialized cells and tissues. Fully differentiated cells, according to Waddington, arise through the canalization of immature cells over time during which the potential for those cells to adopt various cell fates is decreased until finally the cells reside in "valleys" that correspond to the "epigenetic" region where fully developed cells of a particular type are located [1].

As more quantitative tools have been introduced into developmental biology, researchers began to use dynamical systems theory to better characterize Waddington's abstract notion of a landscape [2]. Under this updated paradigm, researchers theorized that the gene expression space (representing the levels of individual mRNAs for each gene) could capture the information present in Waddington's proposed "epigenetic landscape" [2]. Thus, the shape of the landscape could be characterized as the result of a dynamical system that regulate gene expression, with the "valleys" being basins of attraction [3]. This paradigm implies a well-defined density around the centers of these basins, with most cells near the bottom of the "valley" and fewer and fewer cells the further away one goes, as expected in the region of a stable attractor [4].

While these ideas regarding Waddington's Landscape have been postulated for over 60 years, it was only until the recent advent of single cell methods that data could be generated to evaluate the validity of the paradigms based on Waddington's Landscape. By characterizing the local density distribution of single-cell data, we can determine whether the data quantitatively fits the predictions of the landscape picture.

II. Methods and Findings

To test the validity of current paradigms based upon Waddington's Landscape, we used a graph theory-based approach that makes connections (i.e. *edges*) between two cells in the original high-dimensional mRNA space if the distance between them falls within a specified cutoff (termed ε). The number of connections each cell has (called the "degree" of that cell in graph theory) represents the local density of other cells around the cell in question. In other words, if a cell has only one neighbor within this distance threshold, it is in a much lower-density region of the space than a cell that has, say, hundreds of neighbors. As a result, the classic "degree distribution" in graph theory can characterize how this density is distributed across all the cells in a particular dataset [5]. Using this approach, we found:

A. All the datasets we analyzed had approximately scale-free density distributions, with observed densities varying over orders of magnitude.

Regardless of the experimental modality we analyzed, ε-cutoff we chose, or distance metric we used, we observed an approximately scale-free degree distribution. This is completely inconsistent with what we would expect for cells the basin of attraction of a cell type [4].

B. Scale-free-like density frustrates cell type separation and analysis.

We found that this scale-free-like density is not dependent on the type of cell being analyzed. In addition, the low slope of the degree distribution, approximately -1, implies that summary statistics like the "mean" expression vector are likely less meaningful than we would expect, a fact which complicates the interpretation of these data and characterization of differential gene expression [5]. We demonstrate that this fact, combined with the highly heterogenous local neighborhoods observed in our ε-networks, prevent easy separation and classification of cells into discrete types

III. Conclusion

We found that the density distribution observed in single-cell data does not conform from the attractor structure expected in the traditional interpretation of Waddington's Landscape currently thought to explain the molecular basis of development. Our findings have theoretical and practical implications for how heterogeneity is conceptualized in development, tissue homeostasis, and disease.

References

[1] Baedke, J. The epigenetic landscape in the course of time: Conrad Hal Waddington's methodological impact on the life sciences. *Stud. Hist. Philos. Biol. Biomed. Sci.* **44**, 756–773 (2013).

[2] Wang, J., Zhang, K., Xu, L. & Wang, E. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci.* **108**, 8257–8262 (2011).

[3] Ferrell, J. E. Bistability, Bifurcations, and Waddington's Epigenetic Landscape. *Curr. Biol.* **22**, R458–R466 (2012).

[4] Strogatz, S. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. (1994).

[5] Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703 (2009).

[1]Bioinformatics Interdepartmental Program, University of California, Los Angeles.
[2]Department of Integrative Biology and Physiology, University of California, Los Angeles
[3]Institute for Quantitative and Computational Biology, University of California, Los Angeles