Quantitative multivariate analysis of dynamic multicellular morphogenic trajectories

Douglas E. White, *Georgia Institute of Technology*
Jonathan B. Sylvester, *Georgia Institute of Technology*
Thomas J. Levario, *Georgia Institute of Technology*
Hang Lu, *Georgia Institute of Technology*
Jeffrey T. Streelman, *Georgia Institute of Technology*
Todd McDevitt, *Emory University*
Melissa Kemp, *Emory University*

Quantitative Multivariate Analysis of Dynamic Multicellular Morphogenic Trajectories

Douglas E. White[a], **Jonathan B. Sylvester**[b], **Thomas J. Levario**[c], **Hang Lu**[a,b,c,d], **Jeffrey T. Streelman**[b,d], **Todd C. McDevitt**[†,a,b,d], and **Melissa L. Kemp**[†,a,d]

[a]The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA

[b]School of Biology, Georgia Institute of Technology, Atlanta, GA

[c]School of Chemical and Biomolecular Engineering, Atlanta, GA

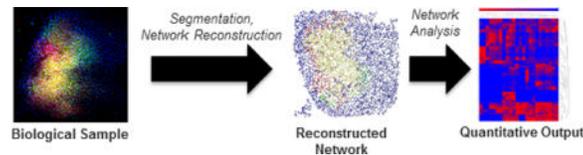[d]The Parker H. Petit Institute of Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA

Abstract

Interrogating fundamental cell biology principles that govern tissue morphogenesis is critical to better understanding of developmental biology and engineering novel multicellular systems. Recently, functional micro-tissues derived from pluripotent embryonic stem cell (ESC) aggregates have provided novel platforms for experimental investigation; however elucidating the factors directing emergent spatial phenotypic patterns remains a significant challenge. Computational modelling techniques offer a unique complementary approach to probe mechanisms regulating morphogenic processes and provide a wealth of spatio-temporal data, but quantitative analysis of simulations and comparison to experimental data is extremely difficult. Quantitative descriptions of spatial phenomena across multiple systems and scales would enable unprecedented comparisons of computational simulations with experimental systems, thereby leveraging the inherent power of computational methods to interrogate the mechanisms governing emergent properties of multicellular biology. To address these challenges, we developed a portable pattern recognition pipeline consisting of: the conversion of cellular images into networks, extraction of novel features via network analysis, and generation of morphogenic trajectories. This novel methodology enabled the quantitative description of morphogenic pattern trajectories that could be compared across diverse systems: computational modelling of multicellular structures, differentiation of stem cell aggregates, and gastrulation of cichlid fish. Moreover, this method identified novel spatio-temporal features associated with different stages of embryo gastrulation, and elucidated a complex paracrine mechanism capable of explaining spatiotemporal pattern kinetic differences in ESC aggregates of different sizes.

Graphical abstract

Segmentation, Network Reconstruction → Network Analysis

Biological Sample → Reconstructed Network → Quantitative Output

Introduction

Model organisms, such as *C. elegans*, *Drosophila*, and zebrafish, are frequently used to interrogate the complex sets of regulatory cues and gene regulatory networks governing morphogenic processes, like gastrulation and neurulation, due to the technical ease in manipulating and imaging [1–3] these organisms. Pluripotent embryonic stem cell (ESC) aggregates present a complementary, alternative *in vitro* platform for investigating mechanisms of morphogenesis due to their intrinsic ability to differentiate into tissues from all three germ layers and yield formation of a variety of primitive tissues including optic cups [4], human intestinal lining [5] and cerebral organoids [6]. Each of these multicellular systems is highly complex both in terms of the heterotypic cell types that comprise the tissue and the emergent spatiotemporal organization dynamics exhibited by heterogeneous cell populations.

Computational modeling of embryonic development has become an increasingly powerful tool to complement experimental investigations due largely to the fact that increased processing speed has reduced the barrier to multiscale simulations of complex multicellular organismal systems. Computational models have been constructed for *C. elegans* [7] and *Drosophila* [8–10] to comprehend the relationships between cell signaling and lineage development in order to gain new insights into the intricate interplay of mechanisms governing development. Stage-specific models have also been developed to examine phenomena such as gastrulation [11] and somite formation [12] at a mechanical level, while computational models of the formation and differentiation of cells in the early mouse embryo [13] and in mouse ESC aggregates [14] explore mechanisms governing early cell fate decisions. Overall, these modeling approaches have provided a wealth of quantitative data to describe spatio-temporal events associated with morphogenesis; however it remains extremely challenging to relate spatial modeling predictions directly with experimental outcomes due primarily to the difficulty in quantifying multicellular pattern features.

This challenge of relating patterns across systems has hindered high-throughput analysis of developmental processes. In experimental systems, divergent phenotypes are often characterized largely by visual inspection, thus lacking the quantitative rigor and objective criteria necessary for direct comparison with computational models. Though several techniques exist to automatically distinguish phenotypes at various spatial scales [15–19], they often lack the resolution of single-cell regulatory dynamics [18, 19], or are customized specifically for investigation of only specific systems [17, 19]. As a result, quantitative metrics extracted from such studies cannot be easily translated between different modes of data analysis or across various model organisms. Hence, a portable pattern recognition pipeline capable of handling various biological and computational inputs would enable direct

quantitative comparisons on previously unattainable spatial and temporal scales between different multicellular systems.

In this report, we demonstrate that deriving physiologically meaningful quantitative metrics capable of distinguishing subtleties between spatial phenotypes in a computationally tractable manner can be achieved via spatially derived networks. This novel approach extracts global metrics, such as path lengths and connectivity information, as well as local metrics based on attributes of specific clusters of cell phenotypes. Identification and comparisons of sub-networks within morphogenic systems allow for greater network quantification and significantly enriches the possible metric space by extracting specific subpopulation information. These metrics can be subsequently analyzed with standard classification methodologies and high dimensional data visualization techniques to extract biologically relevant information[20–22]. Furthermore, this network-based scheme is the first classification approach currently capable of using spatial data at the single cell level to classify tissue level pattern dynamics. To demonstrate this, we first explored the use of simple network metrics to capture patterning across a large range of parameter space using computational simulations, before applying this analysis to a variety of applications. Several input formats (i.e. 2D images and 3D confocal data sets) were analyzed to demonstrate the robust nature of this technique. Consequently, we describe the creation of a powerful and modular pattern identification algorithm with sufficient portability to address meaningful questions about the spatiotemporal dynamics of biological pattern formation. This methodological tool automatically identified stages of gastrulation in cichlid fish in an unprecedented manner, provided the first quantitative description of spatio-temporal patterns associated with loss of pluripotency in ESC aggregates, and uncovered a paracrine mechanism capable of explaining the observed differences in spatiotemporal pattern kinetics associated with ESC aggregate differentiation.

Results

Network-based metric identification for spatial pattern classification

In order to identify appropriate metrics for characterizing spatial patterns, a flexible and expedient computational framework was necessary to generate a large data set comprised of various types of multicellular patterns. We generated an *in silico* training set by using pattern generation algorithms to simulate spatial features of differentiation in 3D multicellular aggregates [14]. In total, seven different classes of general patterns were generated: random, outside-in, inside-out, snaked, globular, undifferentiated (>95% Oct4+) and differentiated (<5% Oct4+) (Fig 1A). We postulated that by applying a network-based approach, simple sets of quantitative metrics could be extracted from complex patterns. Typically, metrics extracted from intact networks include path lengths, average connection counts, and connection lengths. While these metrics are important for describing the network as a whole, they did not accurately describe the variations present in biological patterns. Therefore, clusters of cells of the same phenotypes were identified to partition the single larger network into a series of sub-networks in which various metrics related to the number, size and distribution of these sub-networks were defined (Supplementary Fig 1). In the context of stem cell differentiation, two sub-networks (Oct4+ and Oct4-) were examined

because of the transitions from high to low Oct4 protein levels as a result of differentiation and loss of pluripotency[23,24].

To examine the overall shape of the multi-dimensional data set, a dimensional reduction technique, principal component analysis (PCA), was performed. PCA generated a 2D projection of all samples based upon their network features. The scatter plot revealed some overlap between different pattern classes, particularly for the globular and snaked patterns (Fig 1A), but in general the different classes could be readily distinguished from one another. More importantly, a continuum of patterns was observed (Fig 1A), suggesting a natural pattern evolution. The model was able to explain 77.7% of the variance in the data set with three principal components (PCs): 46.5% with PC-1, 13.9% with PC-2 and 11.7% with PC-3 (Fig 1B). PC-1 represented the stage of differentiation and was positively correlated with variables associated with differentiated clusters or overall differentiation, whereas PC-2 and PC-3 were associated with spatial sub-network descriptors. PCA also revealed that each metric contributed significantly to the model, indicating that inclusion of all metrics was important to comprehensively describe the data set.

To verify that simple network-based metrics could describe complex spatial pattern classes, various classification algorithms were assessed to determine their ability to distinguish between pattern types: k nearest neighbor (KNN), state vector machines (SVM, NuSVC), stochastic gradient descent (SGD) and decision tree algorithms. Classification labeling overlapped with the true labels as assessed via PCA dimensional reduction (Supplementary Fig 2). The SVC classifier had the highest overall accuracy (.998), precision (.999), and recall scores (.987), suggesting it was the most appropriate algorithm for classifying spatial patterns. Collectively, these results suggested that novel network-based measurements provide a robust set of quantitative metrics for classification of complex spatial patterns.

Quantitative pattern trajectories of 3D pluripotent stem cell aggregates

Images of Oct 4+ to Oct4– transitions during differentiation of 3D murine ESC aggregates were acquired experimentally to determine how well network metrics captured an *in vitro* dynamic biological process (Fig 2A). Since pluripotent differentiation is known to be modulated by aggregate size[25–27], two starting cell densities were examined (250 and 1000 cells/aggregate). To evaluate the previously derived network metrics for pattern analysis, experimentally-obtained confocal images were converted into a network representation of the cells using a digital reconstruction pipeline (Supplementary Fig 3). A representative time course of differentiation in 1000-cell aggregates (Supplementary Fig 4) demonstrates the fidelity of this process in accurately converting images into annotated networks.

Next, digitized ESC aggregate networks were analyzed with the aforementioned network metrics. PCA revealed an average trajectory through latent variable space for cell aggregates in which all cells began in an undifferentiated state and proceeded through a transitioning period until finally settling into a differentiated state (Fig 2B). Regardless of size, the trajectories substantially overlapped, although the 250-cell trajectory was more truncated temporally relative to the 1000 cell aggregates, reflecting accelerated differentiation (Supplementary Fig 5). The PCA model explained 76.1% of the variance in the data: 43.6% from PC-1, 22.2% from PC-2, and 10.3% from PC-3 (Fig 2C). All metrics significantly

contributed to at least one principal component, verifying that network-derived metrics capture the variance of biological spatial patterns. PC-1 represented differentiation, while PC-2 and PC-3 again correlated with spatial sub-network descriptors representing inter-pattern variation. The principal component metric weights for the experimental data closely mirrored the weights for the *in silico* training set, indicating that network-derived metrics comprehensively capture the inherent biological variance that transpires during the course of ESC aggregate differentiation.

While the Oct4+ and Oct4− states were quite distinct, the intermediate transitioning period displayed a great amount of variance (Fig 2B). To determine if the variation was due to differences in spatial patterns, classification was performed using the previously trained SVC classifier to characterize the distribution of spatial patterns in each state. Classification indicated that the initial state consisted largely of a mix of undifferentiated, random, and outside-in patterns, while the final state consisted of a mix of entirely differentiated and outside-in patterns (Fig 2D, Fig 2E). The variation with respect to each time-point peaked at days 5 and 6 (Fig 2B), which also displayed a diverse set of spatial patterns (Fig 2E). Furthermore, SVC classification predicted that many of the aggregates at day 5 and 6 could belong to multiple pattern classes, indicating that these patterns were more spatially complex and thus displayed components of multiple different pattern types (Fig 2F). Overall, spatial pattern evolution progressed in the following temporal order: undifferentiated, snaked, random, globular, inside-out, differentiated. Similar pattern trajectories were also observed for the 250-cell aggregates (Supplementary Fig 5). These results represent a biological trajectory describing spatial pattern evolution in a portable quantitative fashion, but even more importantly, the analysis suggests that early differentiation in ESC aggregates progresses via quantifiable spatial patterns that do not display purely random characteristics. Furthermore, this is the first description of biological trajectories using single cell information to capture spatial pattern complexity.

Pattern trajectories of computational models describing ESC differentiation

Next, to probe the mechanisms governing the formation of spatial patterns associated with differentiation, an agent-based modeling approach was employed in which cells are allowed to proliferate, migrate, and differentiate within a 3D aggregate configuration[14]. In our prior work, a simple set of rules based on local neighboring cell state(s) were used to govern changes in cells state; however, we found comparisons between modeling results to be nearly impossible because a quantitative set of descriptors for assessing spatial patterns did not exist. In addition, comparison with experimental data could not be directly accomplished without a validated digitization strategy. These challenges were addressed simultaneously by enabling direct comparison between spatial patterns from computational models and experimental results via the use of these newly defined network metrics.

Seven models with different rule schemes driving differentiation were investigated: random, local positive feedback, local negative feedforward, local competing regulation, paracrine activation, paracrine inhibition, and combined paracrine activation/inhibition (Fig 3A). Diffusion simulations were carried out to understand how the ratio of consumption to production of soluble factors effected gradient propagation prior to paracrine rule creation

(Supplementary Figure 6). Simulations were carried out over a six-day period with different initial aggregate sizes: 250- and 1000-cells/aggregate. Each rule set was simulated with multiple parameters to explore the breadth of pattern trajectory space (Supplementary Table 1). PCA using the metrics described in figures 1 and 2, captured 76.5% of the simulation variance: 48.83% from PC-1, 17.7 from PC-2, and 9.9% from PC-3 (Supplementary Figure 7). Again, PC-1 represented differentiation, while the PC-2 was influenced by standard deviations in sub-network measurements, correlating with the formation of spatial patterns in the simulations.

The previously trained pattern classifiers were applied to assess the spatial patterns generated by the computational simulations and analyzed using hierarchical clustering (Figure 3B). While the competing, negative feedforward, and positive feedback rules all generated similar pattern distributions and trajectories, the paracrine rules generated a more diverse set of pattern types. All together the rules were able to achieve a wide variety of diverse pattern types and evolutions. These results indicate that various complex pattern evolutions and temporal kinetics can be achieved using parsimonious, generalized rule sets.

Comparing computational and experimental pattern trajectories

A powerful feature of the employed network-based methodology is the ability to directly compare results across different platforms, thus allowing the modelling and experimental data sets to be merged into a single metric set. PCA was used to assess which metric axes were most important for describing the variation, resulting in a set of 5 axes responsible for ~ 90% of the variance along which to compare the different data sets (Supplementary Figure 6). Previously it was postulated that a competing regulation scheme could capture the spatial pattern evolution during differentiation, but this mechanism failed to explain kinetic differences between 250 and 1000-cell differentiation trajectories. To identify parameter sets and rules that did modulate differentiation based on aggregate size, a ratio of the differentiation rate of 250-cell to 1000-cell differentiation was calculated (Figure 3C). This ratio confirmed that local feedback rules did not exhibit significant size dependent differences, where nearly all of the soluble rules did. Both the paracrine activation and competing paracrine rules resulted in slower differentiation of 1000-cell aggregates than 250 cell aggregates, matching experimental observations. By comparing these rules to the experimental data on the PCA axes derived previously, it was determined that the paracrine competing rule set yielded the best fit (Figure 3F) because this rule accurately captured both the relevant time scales for differentiation (~24 hours in 250 cells/aggregate and ~48 hours for 1000 cells/aggregate) and the spatial pattern evolution (Figure 3D). Furthermore, this rule suggested that in 250 cell aggregates, differentiation was primarily induced by the absence of activator for the pluripotent state, while differentiation was largely caused by the accumulation of an activating factor within the 1000 cell aggregates (Figure 3E). Collectively, these results demonstrate a non-intuitive paracrine mechanism that can accurately explain differentiation of ESC aggregates and thereby demonstrate the power and utility of network based metrics for elucidating new mechanisms governing biological processes.

Network-based analysis of cichlid gastrulation

Finally, to assess the applicability of network-based analysis on a tightly regulated biological process involving multiple biological signals, gastrulation of East African cichlid fish embryos was analyzed. Gastrulation begins from a relatively undifferentiated cell aggregate that undergoes coordinated multicellular movement and differentiation to yield three tissue layers, a neural plate and rudimentary gut[28]. This fundamental developmental process occurs under the tight spatial and temporal control of morphogens, such as bone morphogenic protein (BMP), and subsequent activation of downstream SMAD signaling via phosphorylation (designated as pSmad). BMP signaling is initially expressed across the entire embryo during gastrulation but clears dorsally to form a pSmad gradient[28]. The subsequent amount and rate of BMP removal correlates with expression of the homeobox gene, *dlx3b*[29]. As with pSmad activity, expression of the *dlx3b* gene goes from being ubiquitous across the majority of the embryo to specific and strong expression in the neural plate boundary (Figure 4A). Assessing the temporal and spatial patterns of multiple correlated signals during morphogenic processes represents a powerful new application of network analysis.

The inherent 3D structure of cichlids presented a challenge, thus a 3D segmentation algorithm was implemented to output annotated spatial networks (Supplementary Fig 8). Using the previously defined metrics, three separate sub-networks of cells were analyzed: pSmad+, *dlx3b*+ and pSmad+/*dlx3b* (Fig 4A). In addition, several new pattern classification metrics were added, such as the ratio of pSmad+/*dlx3b*+ nodes to *dlx3b*+ nodes and pSmad+ nodes, as well as circularity and eccentricity measures for cell clusters in an effort to capture the additional spatial complexity of this system. To identify relevant metrics feature extraction analysis was performed (Supplementary Figure 9).

Initial hierarchical clustering analysis of the resulting metrics revealed segregation of the data set into three main clusters, but the majority of the data set fell into a single large cluster that made subsequent interpretation difficult (Fig 4B). Thus, in order to analyze the clusters further, a PCA model was created that explained the majority (83.8%) of the variance: 51.3% from PC-1, 22.9% from PC-2, and 9.6% from PC3 (Fig 5D). PC-1 correlated highly with metrics associated with the shape of pSmad+ and pSmad+/*dlx3b*+ clusters, while PC-2 was strongly inversely correlated with *dlx3b*+ cluster metrics, and PC-3 correlated with eccentricities metrics.

This final PCA model not only revealed the initial and terminal states detected by hierarchical clustering, but more importantly, resolved the remaining data along a clear temporal trajectory (Fig 4C). Selecting various points along the trajectory revealed a set of patterns that matched the known biology, while also identifying subtle transition states between discrete time points (Fig 4E). Early development time points (0 hours – 4 hours) were characterized by a shrinking pSmad+ region with an increase in *dlx3b*+/pSmad+ regions, as indicated by the shift primarily in early time points along PC-2. The midpoint of gastrulation (~4 hours) exhibited an important switch in the formation and shape of the *dlx3b*+ region, and the final developmental stage (4 hours – 8 hours) was heavily influenced by the emergence of a crescent of *dlx3b* expression, as indicated by its progression along the

PC-1 axis. To test how well these metrics predicted the evolution, a set of path finding simulations intended to find the most likely flow of information through a given process were performed (Figure 5A). The average trajectory for these simulations indicated the 0 hours samples peaking first, followed by a peak in the 4 hour, and 8 hour samples (Figure 5B). Analysis of the clusters indicated that early samples were marked by a high level of pSmad+ and pSmad+/*dlx3b*+ regions, followed by a gradual increase in the presence of solely *dlx3b*+ clusters (Figure 5C). Taken together, these results indicate that the biological trajectory produced by this approach can distinguish the precise state of gastrulation of a biological sample regardless of the experimental timepoint at which it was acquired during the process (Figure 4C, 4E, Figure 5B), further demonstrating the unique strength of a quantitative network-based pattern classification approach for more accurately analyzing morphogenic processes.

Discussion

Network and information theory provide a powerful tool for the analysis of many complex systems ranging from social[30,31] to biological networks[32,33]. For the first time, our work applied the principles of network theory to the study of morphogenic biological systems in a spatial manner. Increasingly, examples of emergent spatial patterns are being reported from initial pluripotent states, leading to organoids such as optic cups, cerebral tissues, or others (4–6), however, quantitative descriptions of multicellular patterns are still lacking. For example Warmflash *et al*. recently used radial distance to delineate organization of differentiated phenotypes within micropatterned ESC colonies [4–6, 34], while Herberg *et al*. used a similar method to compare spatial distributions of proteins in ESC colonies to computational models [35], but these approaches are not sufficient to describe more complex 3D structures. Our method reconstructs cellular locations as interacting networks that can subsequently be further subdivided into biologically relevant sub-networks. This network-based approach circumvents problems associated with traditional classification methods that rely solely on standardized images [19] and use of individual pixel classification methodologies [36, 37]. While some systems exist for classifying spatial patterns in zebrafish [17], *C. elegans* [15] and *Drosophila* embryos [19,38–40], previous approaches require specifically oriented and annotated images, are specific to the organism of interest, and/or often do not have single cell resolution. In particular, machine learning algorithms developed for positional dependence of patterning in *Drosophila* have not relied on individual cell segmentation for evolution of network connectivity over time. Our quantitative method is the first network based approach capable of integrating single cell spatial positioning and phenotypic state information to quantitatively describe dynamic tissue level patterning. The multiple examples illustrated herein highlight the broad utility of network-based analysis for identification of spatial biological patterns via the formulation of novel pattern metrics. We report the derivation of pattern trajectories associated with several systems: experimentally observed loss of Oct4 in ESC aggregates, computational models of Oct4 loss in ESC aggregates, and gastrulation in cichlids. Novel biological insights gained using our network analysis approach include: 1) differences in ESC aggregate spatiotemporal pattern kinetics can be explained by a combined paracrine signaling methodology, and 2) gastrulation in cichlid fishes can be partitioned into a set of discrete

stages. In the case of ESCs, a large body of literature exists that suggests differentiation is heavily influenced by ESC aggregate size[26,27,41–43]; soluble gp130 proteins have been identified as a paracrine mechanism which modulate of differentiation [44,45] The two paracrine process proposed here can explain these differences (one in a secreted factor is responsible for maintain pluripotency, and the other where more differentiated cells secrete a factor which induces differentiation), and mirrors the known properties of soluble LIF and FGF4 signaling respectively[24,46–48]. Surprisingly but interestingly, the lack of local neighbor-to-neighbor regulation of phenotypic state, as analyzed by our methodology, suggests that transmission of cell state information by intercellular cues, such as Notch, may impact later stages of differentiation than the time period examined here. Furthermore, this analytical approach uniquely enables the first direct quantitative comparison between computational modeling and complex emergent spatiotemporal patterns during multicellular lineage commitment in 3D ESC aggregates.

Methods

D3 murine ESCs were used for all of these experiments. All cichlid embryo experiments were performed with approval by and in compliance with Georgia Tech Institutional Animal Care and Use Committee. Modelling and comparisons were performed using custom python code. For more detailed methods on data acquisition and analysis, please refer to the supplementary methods section.

Conclusions

In summation, this novel pattern classification pipeline permits entirely new forms of quantitative analysis based upon the fundamental interconnectivity of multicellular networks, which could revolutionize the characterization of biologically complex spatiotemporal phenomena.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Notes and references

1. Tomer R, Khairy K, Amat F, Keller PJ. Nat Methods. 2012; 9:755–763. [PubMed: 22660741]

2. Krzic U, Gunther S, Saunders TE, Streichan SJ, Hufnagel L. Nat Methods. 2012; 9:730–733. [PubMed: 22660739]

3. Keller PJ. Methods. 2013; 62:268–278. [PubMed: 23523701]

4. Eiraku M, Takata N, Ishibashi H, Kawada M, Sakakura E, Okuda S, Sekiguchi K, Adachi T, Sasai Y. Nature. 2011; 472:51–56. [PubMed: 21475194]

5. Spence JR, Mayhew CN, Rankin SA, Kuhar MF, Vallance JE, Tolle K, Hoskins EE, Kalinichenko VV, Wells SI, Zorn AM, Shroyer NF, Wells JM. Nature. 2011; 470:105–109. [PubMed: 21151107]

6. Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, Homfray T, Penninger JM, Jackson AP, Knoblich JA. Nature. 2013; 501:373–379. [PubMed: 23995685]

7. Setty Y, Dalfo D, Korta DZ, Hubbard EJ, Kugler H. Development. 2012; 139:47–56. [PubMed: 22147952]

8. Kazemian M, Blatti C, Richards A, McCutchan M, Wakabayashi-Ito N, Hammonds AS, Celniker SE, Kumar S, Wolfe SA, Brodsky MH, Sinha S. PLoS Biol. 2010; 8

9. Mizutani CM, Nie Q, Wan FY, Zhang YT, Vilmos P, Sousa-Neves R, Bier E, Marsh JL, Lander AD. Developmental cell. 2005; 8:915–924. [PubMed: 15935780]

10. Lembong J, Yakoby N, Shvartsman SY. Proc Natl Acad Sci U S A. 2009; 106:3213–3218. [PubMed: 19218441]

11. Tamulonis C, Postma M, Marlow HQ, Magie CR, de Jong J, Kaandorp J. Developmental biology. 2011; 351:217–228. [PubMed: 20977902]

12. Hester SD, Belmonte JM, Gens JS, Clendenon SG, Glazier JA. PLoS Comput Biol. 2011; 7:e1002155. [PubMed: 21998560]

13. Krupinski P, Chickarmane V, Peterson C. Curr Opin Genet Dev. 2012; 22:613–618. [PubMed: 22959149]

14. White DE, Kinney MA, McDevitt TC, Kemp ML. PLoS computational biology. 2013; 9:e1002952. [PubMed: 23516345]

15. Crane MM, Stirman JN, Ou CY, Kurshan PT, Rehg JM, Shen K, Lu H. Nat Methods. 2012; 9:977–980. [PubMed: 22902935]

16. Zacharia E, Bondesson M, Gustafsson JA, Kakadiaris IA. Conf Proc IEEE Eng Med Biol Soc. 2012; 2012:3998–4001. [PubMed: 23366804]

17. Castro-Gonzalez C, Luengo-Oroz MA, Douloquin L, Savy T, Melani C, Desnoulez S, Ledesma-Carbayo MJ, Bourginey P, Peyrieras N, Santos A. Conf Proc IEEE Eng Med Biol Soc. 2010; 2010:5520–5523. [PubMed: 21096468]

18. Trisnadi N, Altinok A, Stathopoulos A, Reeves GT. Methods. 2013; 62:68–78. [PubMed: 23104159]

19. Puniyani K, Faloutsos C, Xing EP. Bioinformatics. 2010; 26:i47–56. [PubMed: 20529936]

20. Vega SL, Liu E, Patel PJ, Kulesa AB, Carlson AL, Ma Y, Becker ML, Moghe PV. J Biomol Screen. 2012; 17:1151–1162. [PubMed: 22811477]

21. Lau KS, Cortez-Retamozo V, Philips SR, Pittet MJ, Lauffenburger DA, Haigis KM. PLoS Biol. 2012; 10:e1001393. [PubMed: 23055830]

22. Naegle KM, Welsch RE, Yaffe MB, White FM, Lauffenburger DA. PLoS Comput Biol. 2011; 7:e1002119. [PubMed: 21799663]

23. Faunes F, Hayward P, Descalzo SM, Chatterjee SS, Balayo T, Trott J, Christoforou A, Ferrer-Vaquer A, Hadjantonakis AK, Dasgupta R, Arias AM. Development. 2013; 140:1171–1183. [PubMed: 23444350]

24. Munoz Descalzo S, Rue P, Garcia-Ojalvo J, Martinez Arias A. Stem Cells. 2012; 30:2683–2691. [PubMed: 22969005]

25. Niebruegge S, Bauwens CL, Peerani R, Thavandiran N, Masse S, Sevaptisidis E, Nanthakumar K, Woodhouse K, Husain M, Kumacheva E, Zandstra PW. Biotechnol Bioeng. 2009; 102:493–507. [PubMed: 18767184]

26. Kinney MA, Saeed R, McDevitt TC. Integr Biol (Camb). 2012; 4:641–650. [PubMed: 22609810]

27. Mohr JC, Zhang J, Azarin SM, Soerens AG, de Pablo JJ, Thomson JA, Lyons GE, Palecek SP, Kamp TJ. Biomaterials. 2010; 31:1885–1893. [PubMed: 19945747]

28. Sanchez-Arrones L, Stern CD, Bovolenta P, Puelles L. Development. 2012; 139:1034–1044. [PubMed: 22318633]

29. Bielen H, Houart C. Dev Cell. 2012; 23:812–822. [PubMed: 23079599]

30. Paul S, O'Malley AJ. J R Stat Soc Ser C Appl Stat. 2013; 62:705–722.

31. Young HP. Proc Natl Acad Sci U S A. 2011; 108(Suppl 4):21285–21291. [PubMed: 22198762]

32. Ruths J, Ruths D. Science. 2014; 343:1373–1376. [PubMed: 24653036]

33. Di Deco J, Gonzalez AM, Diaz J, Mato V, Garcia-Frank D, Alvarez-Linera J, Frank A, Hernandez-Tamames JA. Curr Top Med Chem. 2013; 13:652–662. [PubMed: 23548026]

34. Warmflash A, Sorre B, Etoc F, Siggia ED, Brivanlou AH. Nat Methods. 201410.1038/nmeth.3016

35. Herberg M, Zerjatke T, de Back W, Glauche I, Roeder I. Cytometry A. 201510.1002/cyto.a.22598

36. Bibicu D, Moraru L, Biswas A. J Digit Imaging. 2013; 26:119–128. [PubMed: 22546981]

37. Mian A, Hu Y, Hartley R, Owens R. IEEE Trans Image Process. 2013; 22:5252–5262. [PubMed: 24107936]

38. Cai JX, Wang H, Huang H, Ding C. Bioinformatics. 2014; 15:16–24. [PubMed: 24428894]

39. Busser BW, Taher L, Kim Y, Tansey T, Bloom MJ, Ovcharenko I, Michelson AM. PLoS Genet. 2012; 8:5252–e1002531.

40. Heffel A, Stadler PF, Prohaska SJ, Kauer G, Kuska JP. Proc Int Conf Image Proc. 2008; 1:721–724. [PubMed: 20046820]

41. Bauwens CL, Song H, Thavandiran N, Ungrin M, Masse S, Nanthakumar K, Seguin C, Zandstra PW. Tissue Eng Part A. 2011; 17:1901–1909. [PubMed: 21417693]

42. Hwang YS, Chung BG, Ortmann D, Hattori N, Moeller HC, Khademhosseini A. Proc Natl Acad Sci U S A. 2009; 106:16978–16983. [PubMed: 19805103]

43. Bauwens CL, Peerani R, Niebruegge S, Woodhouse KA, Kumacheva E, Husain M, Zandstra PW. Stem Cells. 2008; 26:2300–2310. [PubMed: 18583540]

44. Davey RE, Nishi K, Mahdavi A, Zandstra PW. FASEB J. 2007; 21:2020–32. [PubMed: 17356004]

45. Davey RE, Zandstra PW. Stem Cells. 2006; 24:2538–48. [PubMed: 16825607]

46. Herberg M, Kalkan T, Glauche I, Smith A, Roeder I. PLoS One. 2014; 9:e92496. [PubMed: 24643025]

47. Ohnishi Y, Huber W, Tsumura A, Kang M, Xenopoulos P, Kurimoto K, Oles AK, Arauzo-Bravo MJ, Saitou M, Hadjantonakis AK, Hiiragi T. Nat Cell Biol. 2014; 16:27–37. [PubMed: 24292013]

48. Dowell KG, Simons AK, Bai H, Kell B, Wang ZZ, Yun K, Hibbs MA. Stem Cells. 2014; 32:1161–1172. [PubMed: 24307629]

Insight, innovation, integration

Decades of research have focused largely on characterizing individual cellular features, and consequently the image processing tools used to analyze biological systems have focused on identifying and quantifying cells as independent, static entities rather than interconnected, dynamic systems. Our *innovation* in analyzing multicellular systems arises from characterizing metrics from dynamic cell-cell networks that allows cross-comparison from virtually any modality of imaging. Because our analysis is platform independent, quantitative *integration* between agent-based computational models and experimental data is feasible. This approach allowed us to gain new *insight* in the pattern evolution of differentiation within stem cell aggregates as dictated by competing paracrine signaling mechanisms.
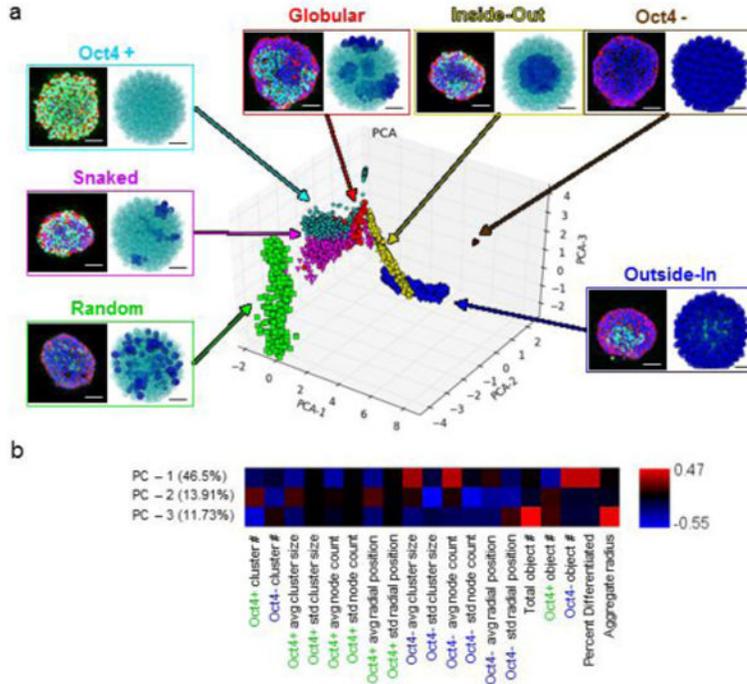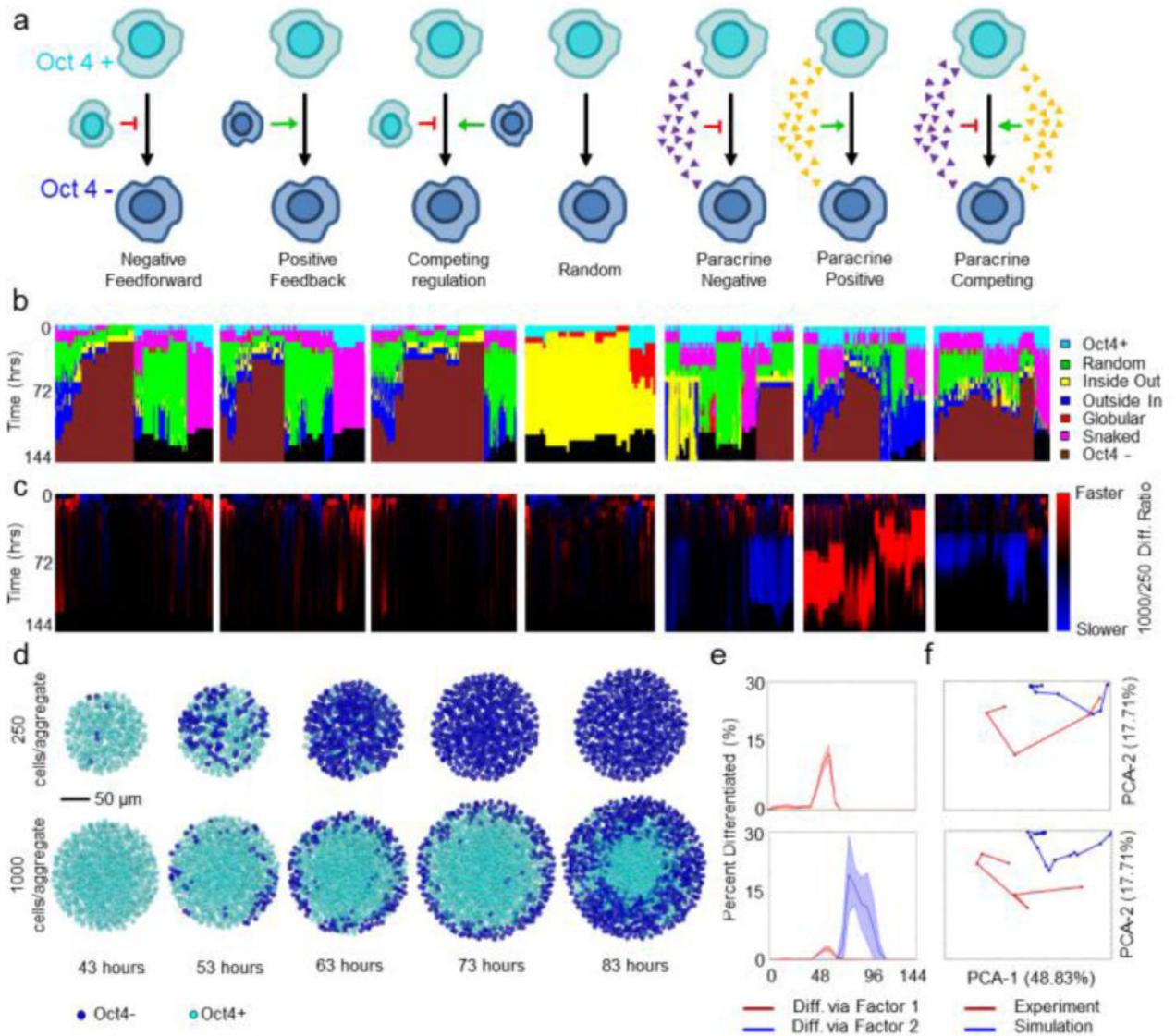
Figure 1.
Validation of spatial network metrics for pattern classification. **(a)** Examples of experimental (green = Oct4 Ab, blue = DAPI, red = phalloidin) and *in silico* generated pattern classes (teal = Oct4+, blue = Oct4 −). Scale bars = 100 μm and 35 μm for the experimental and *in silico* images, respectively. Representative images are mapped onto the PCA plot of the resulting pattern space, color coded by individual pattern type. **(b)** PCA axis analysis indicating the relative contribution of each metric to the given principal component axis.
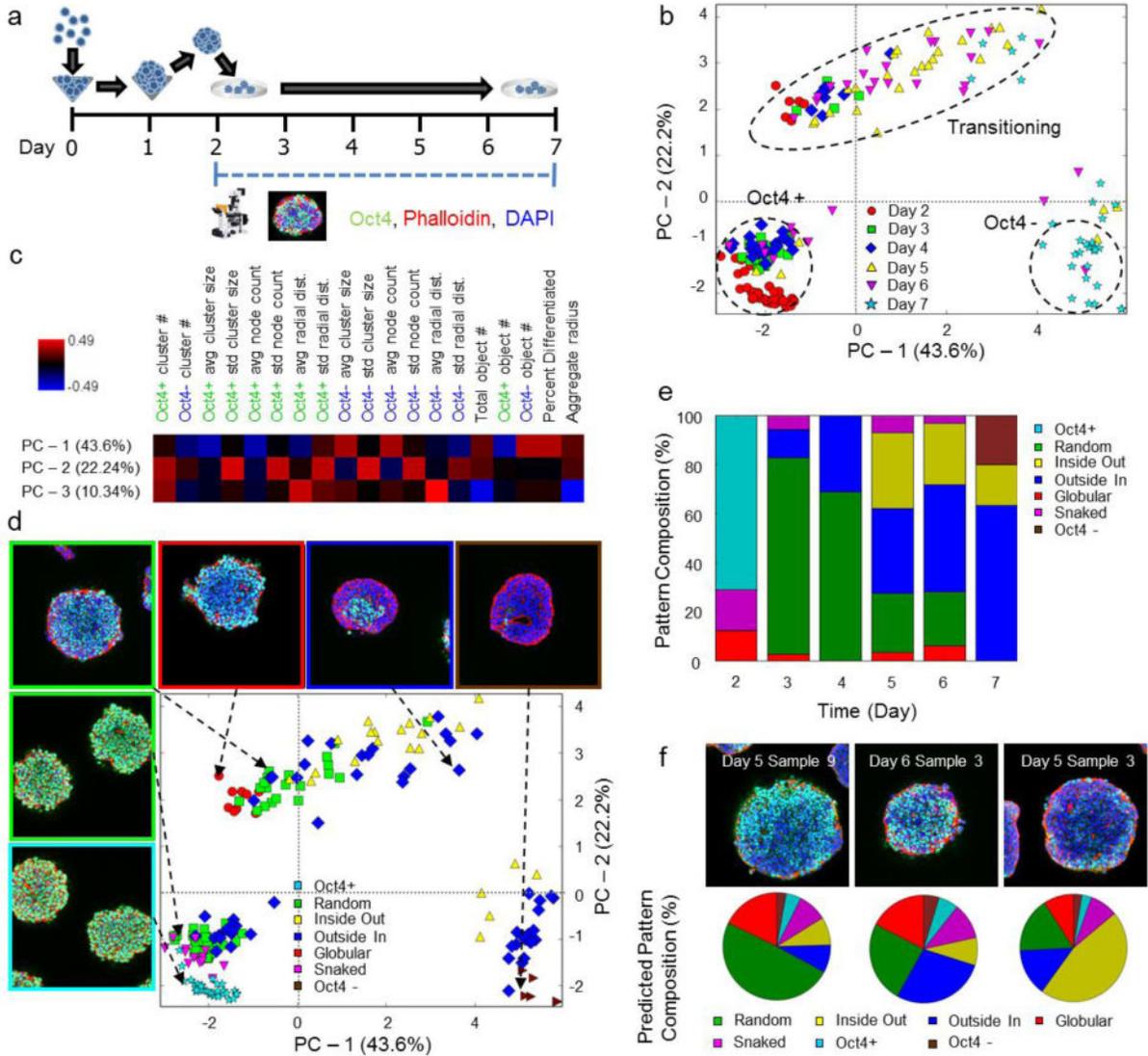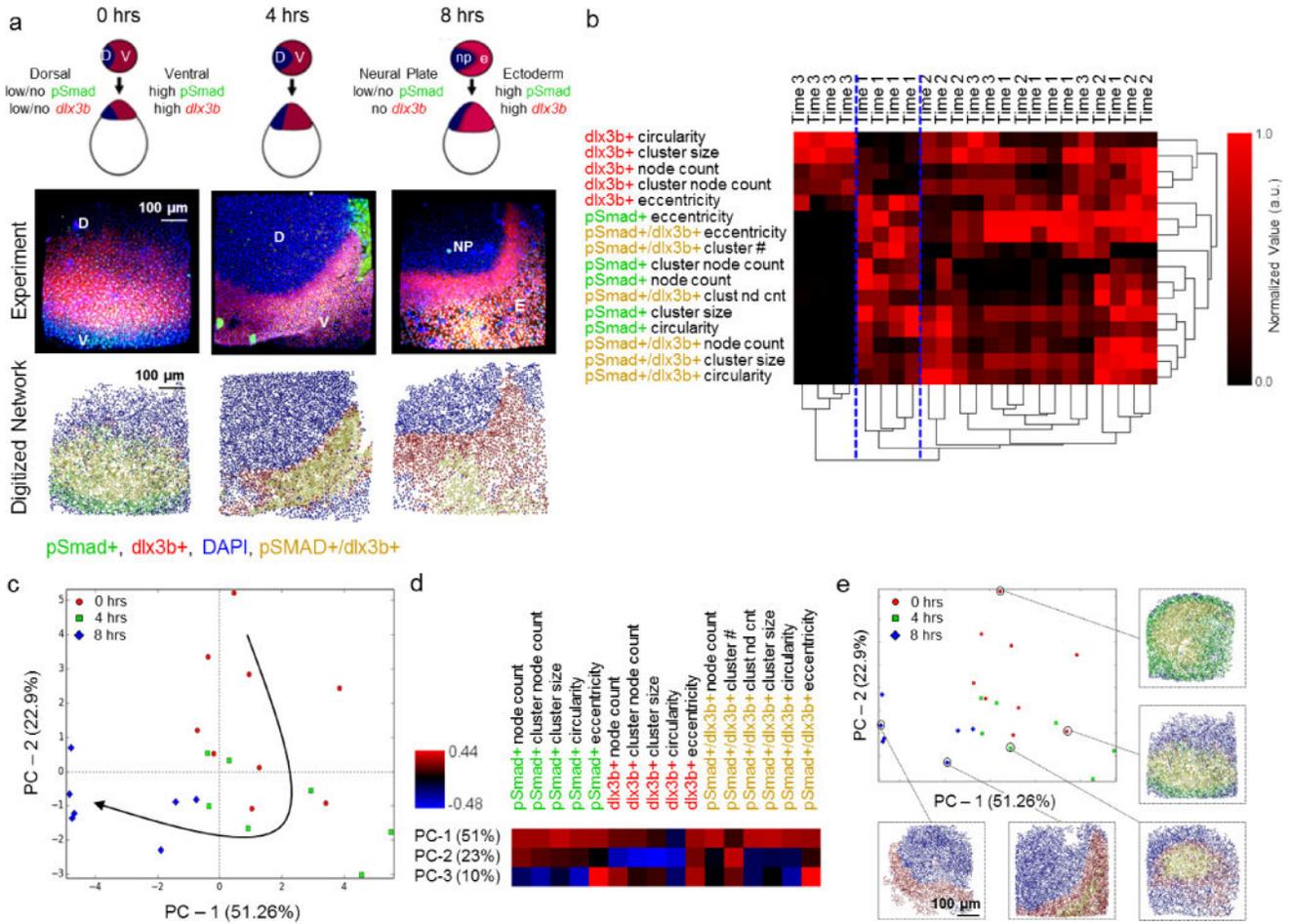
Figure 2.
Evaluation of spatial and temporal patterns during early ESC aggregate differentiation. **(a)** Experimental plan and time course; confocal microscopy was performed daily (on days 2–7). **(b)** PCA trajectory for 1000 cell aggregates. Dashed regions highlight different cell states present. **(c)** Heat map of the weights of each metric in relation to each principal component. **(d)** Annotations for the pattern classification of each sample. **(e)** Pattern compositions for each day of differentiation. **(f)** SVC classification of samples that contain aspects of multiple core pattern classes. Experimental images (top row) and, predicted pattern composition (pie charts, bottom row). The predicted pattern composition is represented as a percent of the different sub-patterns contained within.
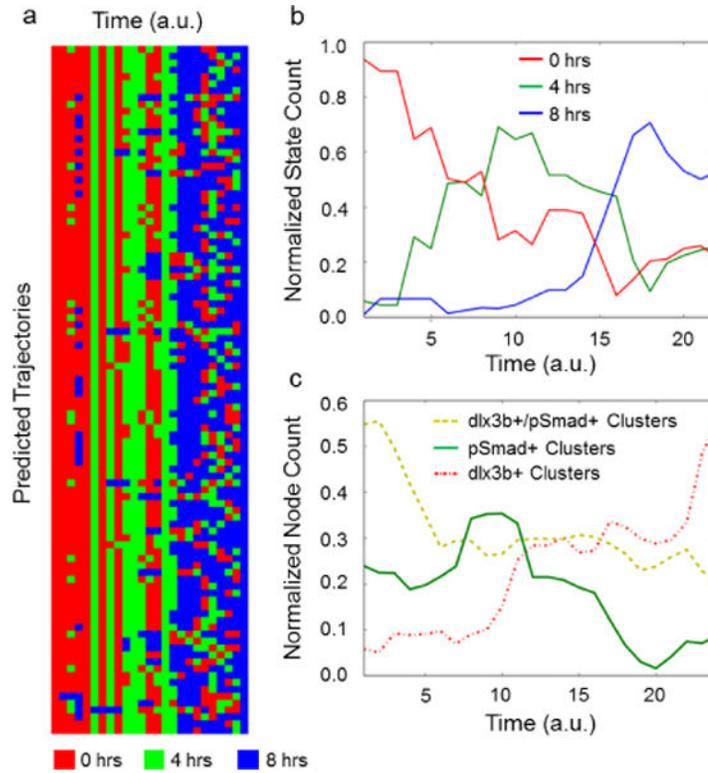
Figure 3.
Characterization of computational pattern trajectories and comparison to experimental patterns. **(a)** Description of rules governing modeling. Red arrows, negative feedback; green arrows, positive feedback; small triangles indicate a soluble morphogen gradient. **(b)** Hierarchal clustering using a Ward linkage algorithm of the pattern trajectories for each parameter set, annotated by rule (gray scale bar). **(c)** Ratio of the 1000-cell to 250-cell differentiation rate. Blue represents slower differentiation in 1000-cellular aggregates, while red represents faster differentiation. **(d)** Representative images of the differentiation trajectories predicted by the competing paracrine rule (cells in blue are Oct4−, cells in teal are Oct4+). (e) Percent differentiation for representative 250 (top) and 1000-cell (bottom) simulations. Red represents differentiation induced via absence of inhibitor, while blue represents differentiation induced via the agonist. (f) PCA projections indicating the best fits of the 250 (top) and 1000-cell (bottom) experimental data (red) to the simulations (blue).

Figure 4.
Pattern analysis during gastrulation reveals the importance of the transition from a pSmad+ dominated early gastrulation phase to dlx3b+ crescent formation in late stage gastrulation. (a) Representative schematic of gastrulation in cichlids on top, experimental confocal data in middle (red – dlx3b, green – pSmad, blue DAPI, yellow pSmad+/dlx3b+) followed by network reconstructions on the bottom (red – dlx3b, green – pSmad, blue DAPI, yellow pSmad+/dlx3b). Scale bars are 100 µm. (b) Hierarchical clustering of the data set yields three distinct sub-clusters (highlighted by dividing blue dashed lines). (c) PCA of the resulting data set reveals (n >= 7 per time point) a distinct trajectory informed by time (0 hours – red circle, 4 hours – green square, 8 hours -blue diamond). (d) Heatmap of how each metric contributes to the principal component model. Blue indicates a strong negative contribution, while red indicates a strong positive contribution. (e) Annotated PCA plot showing the evolution of pattern formation over time during gastrulation in cichlid fish. All images have the same scale bar of 100 µm.

Figure 5.
Information analysis predicts evolution of images during gastrulation. (**a**) Predicted pattern trajectories through time exhibit correct temporal ordering of early (0 hours, red), middle (4 hours green) and late gastrulation (8 hours blue). (**b**) The averages for all trajectories reflect state evolution from early to middle to late gastrulation. (**c**) The averages for all trajectories indicate the evolution of different cluster number during gastrulation.