Evolution of Cellular Heterogeneity in a Nascent Multicellular Organism by Image-based Single-Cell Profiling

Kai Tong Aditya Milind Vadhavkar Jayant Prakash

***Abstract*—The origin of multicellular organisms from unicellular ancestors is one of the major evolutionary transitions in the history of life. Cell differentiation is a hallmark of multicellularity, but it remains elusive how nascent multicellularity evolves cell differentiation. The multicellular snowflake yeast is experimentally evolved from the unicellular yeast *Saccharomyces cerevisiae* under selection for larger size, and presents a unique model system for characterizing, in real time, how cellular heterogeneity evolves in a nascent multicellular organism, which may pave the way for the evolution of cell differentiation. Here we established a pipeline to extract eight image-based features of single cells within snowflake yeast colonies, and used it to measure and compare the cellular heterogeneity across four evolutionary time points, namely, week 0, week 8, week 16 and week 24. To identify potential cell states/types within each week, we applied t-SNE and DBSCAN with stringent criteria for determining the number of cell clusters. In week 0, we only found one cell cluster, while in all other weeks, we found one dominant cell cluster and a few other tiny cell clusters with unknown biological implications. By comparing the largest cell cluster in different weeks, we found significant evolutionary changes in most cellular features as well as varying evolutionary rates between different weeks. This study established a technical framework for examining the evolution of cellular heterogeneity in a nascent multicellular organism, which can be applied to investigate other evolutionary time points, other evolutionary lineages, and other image-based cellular features.**

I. Introduction

Evolution of multicellular organisms from unicellular ancestors is one of the major evolutionary transitions in the history of life, giving rise to macroscopic, extraordinarily complex life forms and forever changing the surface of the earth (1, 2). Cell differentiation, the generation of structurally and functionally distinct cell types that work together to build a multicellular body, is one of the hallmarks of multicellularity (3). However, it remains elusive how simple, undifferentiated nascent multicellular organisms evolve cell differentiation. This question may be approached by examining how cell differentiation evolved in extant multicellular lineages (e.g., animals, plants) in history, but this approach is hampered by the ancient origin of multicellularity and the lack of fossil records representing the transitional forms (4).

To circumvent these limitations, we experimentally evolved multicellularity *de novo* for characterizing multicellular evolution in real time. Using settling selection to select for larger sizes that settle faster through liquid medium, we evolved the unicellular Bakers yeast *Saccharomyces cerevisiae* into multicellular colonies (termed snowflake yeast) over months of in-lab evolution (5). This nascent multicellular phenotype is the result of a loss-of-function mutation in the *ACE2* gene, resulting in incomplete cell division and thus clonal multicellular development (6), which makes the cellular heterogeneity within snowflake yeast colonies affected only by non-genetic factors. Therefore, snowflake yeast presents a unique model system for characterizing if and how cells in a nascent multicellular organism may evolve to become different from each other, which may reflect or pave the way for the evolution of cell differentiation.

In this study, we characterized the evolution of cellular heterogeneity within snowflake yeast colonies over 24 weeks of settling selection, based on a set of image-based cellular features and unsupervised learning.

II. Methods

A. Snowflake yeast strains

All snowflake yeast strains used in this study were taken from one lineage in an evolution experiment conducted by Jennifer Pentz in Will Ratcliffs lab. In this experiment, an *ACE2* knockout strain, which is multicellular, experienced settling selection for a year and evolved larger multicellular sizes over time. This study focuses on four evolutionary time points, namely, week 0, week 8, week 16 and week 24, from each of which one clonal isolate was taken to represent the evolving population.

B. Image-based single-cell profiling in snowflake yeast

In order to characterize the cellular heterogeneity within snowflake yeast colonies, we established a pipeline for extracting image-based cellular features (Figure 1). Snowflake yeast was inoculated into 10 ml of YPG medium and grown for 1 day, and then split 1:100 and grow for another 12 hours to stay at exponential phase. The culture was diluted, washed with water, and resuspended in YNB-G medium containing 1% concanavalin A (ConA), a green fluorescent dye that stains yeast cell wall for easier cell segmentation. Sample was pipetted on a microscope slide and the coverslip was pressed gently to crush multicellular colonies into a single cell layer for imaging.

Images of snowflake yeast colonies were taken by automated multi-point acquisition at 40x magnification with oil-immersion lens. For each point, the focal plane was determined by Auto Focus system, and images at brightfield channel and green fluorescent channel were acquired. Following image acquisition, low-quality images (e.g., touching clusters, poor ConA staining, out of focus) were manually excluded.

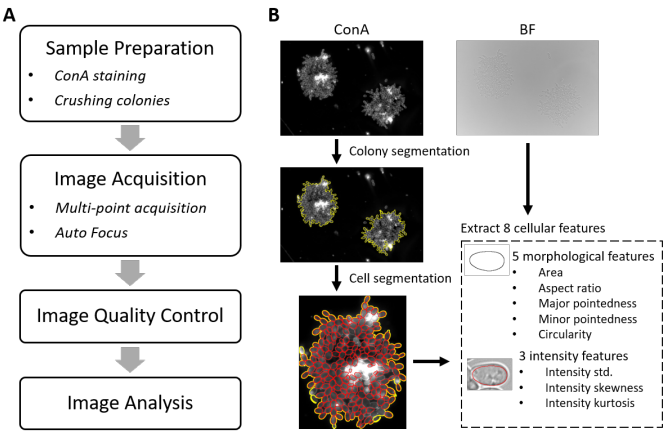

Figure 1. Pipeline for image-based single-cell profiling in snowflake yeast. (A) Overview of the pipeline. (B) Major steps of image analysis. Based on ConA staining, colonies are identified (shown by yellow contours), and cells within each colony are segmented (shown by red contours). Based on cell segmentation results and brightfield (BF) images, eight cellular features are extracted per cell.

Image analysis was performed using custom ImageJ scripts (Figure 1B). Based on ConA staining, colonies were segmented by performing Auto Local Threshold (method mean, radius 20), Close, Fill Holes, Erode, and Analyze Particles, and incorrect colony segmentations were manually removed. Within each detected colony, cells were segmented by performing Auto Local Threshold (method mean, radius 10), Despeckle, Invert, Close, Fill Holes, Watershed for Irregular Features, Shape Smoothing, and Analyze Particles with filtering based on large size, high circularity and high solidity. Based on the cell segmentation results and brightfield images, eight features were extracted for each detected cell to reflect its overall physiological status, including five morphological features and three intensity features. The morphological features of the cell include area, aspect ratio (of the fitted ellipse of the cell), major and minor pointedness, and circularity. Major/minor pointedness were used to account for the deviation of the centroid (C) of the cell from the centroid of the bounding rectangle (CB) of the cell, which is projected onto the major/minor axis and divided by half of the length of major/minor axis (a,b). Major Pointedness=$\frac{|C_a - CB_a|}{a/2}$, Minor Pointedness=$\frac{|C_b - CB_b|}{b/2}$. The intensity features reflect the texture of cytoplasm, and include the standard deviation, skewness, and kurtosis of the intensity values of all pixels within the segmentation contour of the cell.

C. Dimensionality reduction, clustering and differential analysis

Initially we did some exploratory data analysis (EDA) by plotting each of the 8 features with other features and observed how the data looks like. We also plotted data for a particular week vs data for all other weeks to see if we could identify any trends as time progressed. After EDA we applied different dimensionality reduction techniques namely Principal Component Analysis (PCA), Independent Component Analysis (ICA), Randomized Projections (RP) and T-distributed Stochastic Neighbor Embedding (t-SNE) and plotted the results obtained from each of these algorithms and then we fed in the dimensionally reduced data matrices to different clustering algorithms namely K Means clustering (KMeans), Expectation Maximization (EM), Spectral Clustering and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). For KMeans and EM we plotted graphs for Silhouette score vs number of selected clusters and Bayesian Information Criterion (BIC) vs number of selected clusters. We selected that value of number of clusters which corresponded to the highest silhouette score and lowest BIC. Although we were able to identify different clusters using KMeans and EM the results were not biologically reasonable. KMeans and EM just divided a single cluster into different smaller clusters which were not visually or biologically different.

We reduced the data to two dimensions using t-SNE and then fed the dimensionally reduced data to DBSCAN. To find the most biologically significant value of number of clusters, we plotted the graph between eps and number of cells and selected the value of number of clusters using the elbow method. In addition to this method we also considered some assumptions in order to decide on the final number of identified clusters. The first assumption was that for each week there should be one big cluster and a few small clusters as we are able to see a similar pattern using EDA and using microscopic analysis of cells in the lab. The second assumption was, for each week we identified different clusters and compared those clusters using the 8 features by applying t test between cells belonging to a particular cluster within that week and all other cells not belonging to that cluster within that week and identified all those features whose p values were below 0.01. We said that a cluster is different from other clusters only if it had 2 or more features that had p values less than 0.01. From this analysis we were able to finalize the number of clusters in each week and also identify the features contributing most significantly towards distinguishing the clusters from one another. Apart from comparing clusters within each week to one another we also performed an analysis wherein we compared the largest clusters identified in each week with one another by applying t test on each of the 8 features. So, all cells belonging to the largest cluster in week 0 were compared to the cells belonging to the largest cluster in week 8 and so on. From this analysis we were able to identify those features that changed the most over time and we were able to comment on the rate of evolution between different time points. t-SNE along with DBSCAN produced the most visually and biologically distinct clusters and thus we decided to go forward with these algorithms.

III. Results

Using the pipeline for image-based single-cell profiling in snowflake yeast, we measured 2524 cells from 14 colonies in week 0, 4001 cells from 15 colonies in week 8, 1167 cells from 7 colonies in week 16, and 4912 cells from 23 colonies in week 24. For each cell, we extracted eight cellular features.

Using the methods described above we found that week 0 had only a single cluster, whereas week 8, week 16 and week 24 had 3 clusters each (Figure 2A-D). From the differential analysis that we performed on the clusters within each week (Figure 2A-D), we found that the clusters identified within a particular week were quite distinct from those identified within other weeks.

Moreover, the largest clusters in each week differed from the largest clusters in all other weeks. Another interesting trend that we observed was that the evolutionary changes from week 0 to week 8 and week 16 to week 24 was larger than that from week 8 to week 16 (Figure 3A). We also found that cellular area and aspect ratio appeared most frequently among the statistically significant features that were able to distinguish between different clusters within weeks as well as across weeks. Thus we decided to analyze these features in detail and when we plotted the graphs for Area vs Aspect Ratio for all weeks, we found that week 0 had a bimodal distribution of cells which was unexpected (Figure 3B). On further analysis in lab and repeating the experiment we suspect that this phenomenon could be caused by a certain bistable gene expression in yeast when cultured in the YPG medium. In week 8 we observed that Area and Aspect ratio had increased significantly from week 0 (Figure 3C). In week 16 Area remained approximately the same but Aspect Ratio decreased slightly from week 8 (Figure 3D). In week 24 Area increased significantly but Aspect Ratio decreased sharply from week 16 (Figure 3E).These results yielded by computational analysis were also confirmed by cell images in the corresponding weeks (Figure 3B-E).

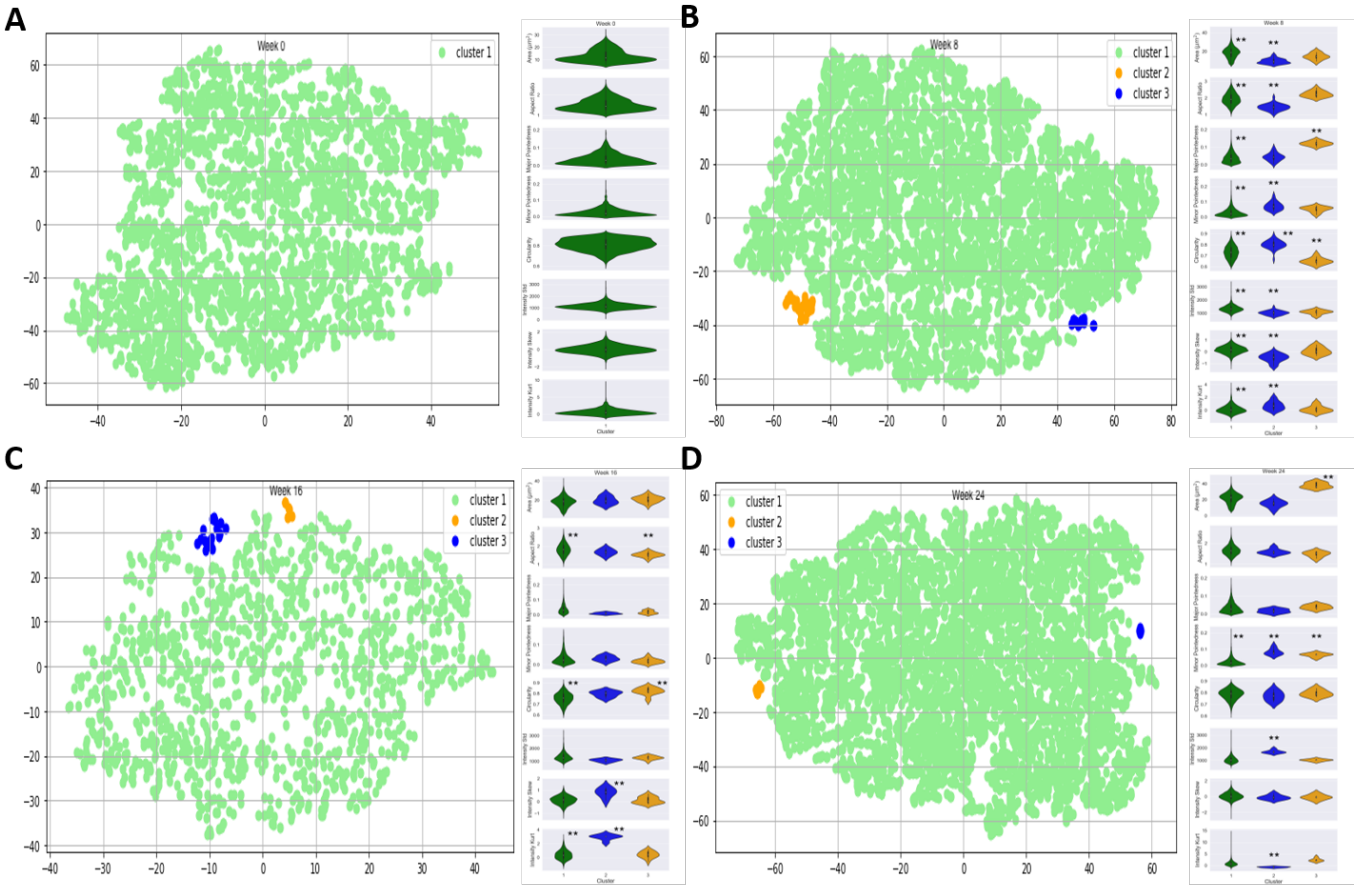

Figure 2. Identification of cell clusters and their markers in week 0 (A), 8 (B), 16 (C) and 24 (D).

IV. Discussion

In this study, we established a pipeline for measuring cellular heterogeneity within snowflake yeast colonies based on a set of image-based cellular features. Using this pipeline, we analyzed four evolutionary time points (week 0, week 8, week 16, week 24) of a snowflake yeast population that had been experiencing settling selection for over a year. We applied tSNE and DBSCAN to identify distinct cell states/types within each week. In week 0, we only found one cell cluster, while in all other weeks, we found one dominant cell cluster and a few other tiny cell clusters with unknown biological meanings that are worthy of further investigations. By comparing the largest

cell cluster in different weeks, we found significant evolutionary changes in most cellular features we studied. Moreover, it appeared that more evolutionary changes happened between week 0 and 8 as well as between week 16 and 24 than between week 8 and 16, reflecting varying evolutionary rates over time.

The current pipeline for image-based single-cell profiling has several limitations. Firstly, ConA often causes local over-staining in some part of a snowflake yeast colony or in some part inside of a cell (Figure 1B), which can lead to incorrect cell segmentation. In this study, we set a high bar to exclude incorrect cell segmentation while still capturing most cells in a colony. In the future, we will optimize the staining protocol and/or the image analysis algorithms to improve cell segmentation. Secondly, crushing multicellular colonies into a single cell layer makes image acquisition easier, but this can also lead to artifacts cells in local dense regions of a colony can be squeezed and slightly deformed, crushing can break small branches from a colony, and there can still be few overlapping cells (i.e., not a perfectly single cell layer). In the future, we will avoid crushing colonies, but resort to confocal imaging and 3D cell segmentation for better measuring the cellular features.

In sum, we established a technical framework for characterizing the evolution of cellular heterogeneity in snowflake yeast, a nascent multicellular organism. This technical framework, with further optimization, can be used to examine other evolutionary time points and other evolutionary lineages, with additional image-based cellular features. Moreover, it also paves the way for investigating how various factors, like colony size, cell age and cell location may contribute to the evolution of cellular heterogeneity.

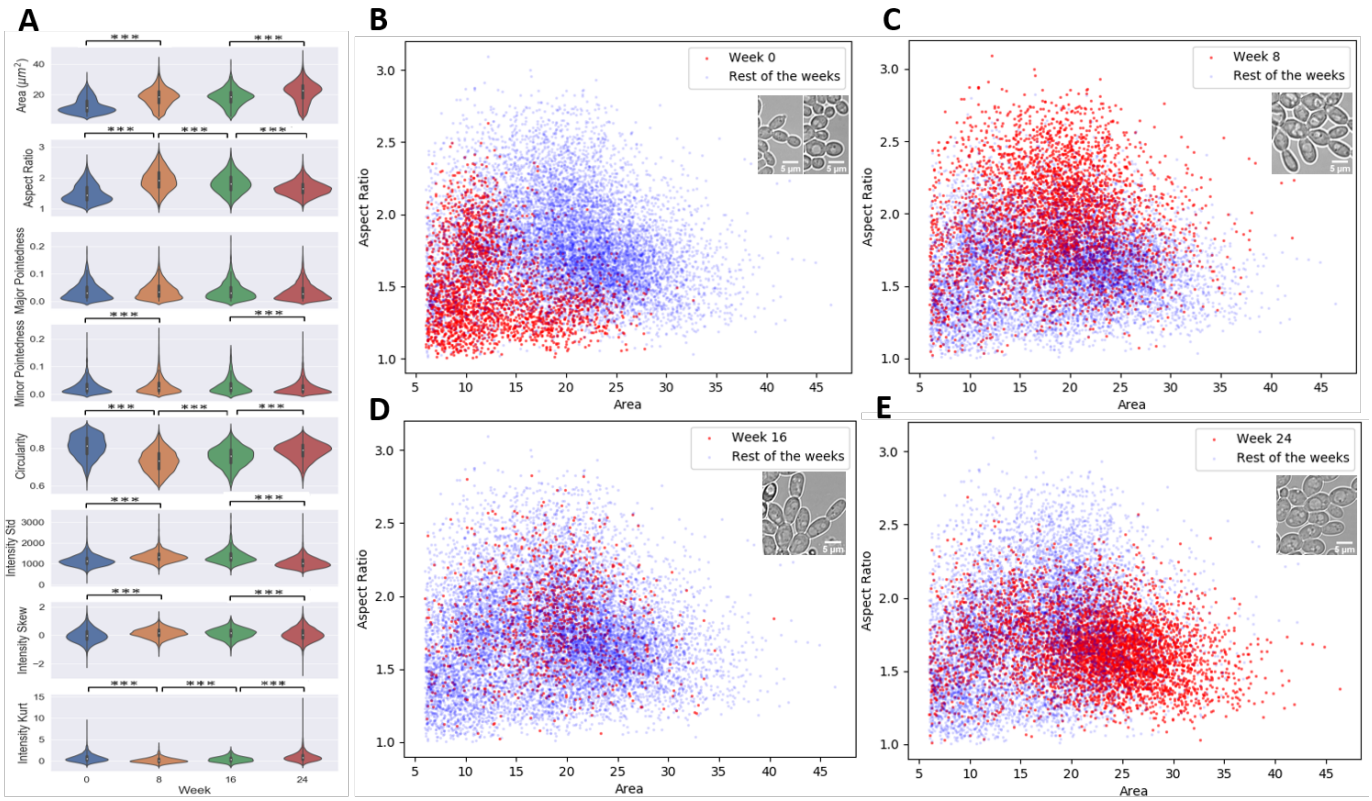

Figure 3. Evolutionary changes across week 0, 8, 16 and 24. (A) Evolutionary changes in all eight cellular features. (B-E) Evolutionary changes in area and aspect ratio, along with their corresponding cell images.

References

[1] West SA, Fisher RM, Gardner A, Kiers ET (2015) Major evolutionary transitions in individuality. Proceedings of the National Academy of Sciences of the United States of America 112(33):10112-10119.

[2] Grosberg RK Strathmann RR (2007) The Evolution of Multicellularity: A Minor Major Transition? Annual Review of Ecology, Evolution, and Systematics 38(1):621-654.

[3] Niklas KJ (2014) The evolutionary-developmental origins of multicellularity. American journal of botany 101(1):6-25.

[4] Knoll AH (2011) The Multiple Origins of Complex Multicellularity. Annual Review of Earth and Planetary Sciences 39(1):217-239.

[5] Ratcliff WC, Denison RF, Borrello M, Travisano M (2012) Experimental evolution of multicellularity. Proceedings of the National Academy of Sciences of the United States of America 109(5):1595-1600.

[6] Ratcliff WC, Fankhauser JD, Rogers DW, Greig D, Travisano M (2015) Origins of multicellular evolvability in snowflake yeast. Nature communications 6:6102.