Multicellular growth as a dynamic network of cells

**Abstract (245 words)**

Cell division without separation leads to the formation of multicellular clusters. The dynamics that allow clusters to grow and later fragment while preserving cellular interactions remain elusive. We examine a dynamic network model, MultiCellNet, with cells as nodes and edges as links between mother and daughter cells. We simulate the growth of the network proportional to the cell division rate (λ) and the network fragments at a rate proportional to the link breaking rate (δ) and inversely proportional to the kissing number, i.e., maximum attainable connections per node (κ). We exclude any free parameters by experimentally controlling all the biological counterparts of model inputs in multicellular yeast clusters. We discover a principle that governs the size of multicellular clusters, in which cell division rate sets a threshold for link breaking rate and kissing number sets the maximum cluster size across all conditions tested. We use this framework to understand how a differentiating multicellular cluster with two cell types, faster-growing germ cells and their somatic derivatives, maintains compositional homeostasis during growth and fragmentation. We find that the ratio of two cell types in these clusters can be experimentally controlled by adjusting the switching and relative growth rates. Moreover, the noise in the composition is minimized by increasing the kissing number. MultiCellNet makes it possible to examine how molecular knobs control the size and cellular composition of the multicellular clusters that were likely the ancestors of more sophisticated forms of multicellular development, organization, and reproduction.

**Introduction**

How unicellular organisms gather to exhibit multicellular behaviour hints at the origins of multicellularity1–32. Selecting for either faster settling4 or nutrient limitations5 leads to the spontaneous evolution of multicellular clusters in *Saccharomyces cerevisiae.* Once multicellular, how these clusters evolved to maintain both size and composition (cell types) remains an open question. The cluster size, i.e., the number of cells per cluster, determines the fitness where geometric cluster size6 or public goods sharing7 provides a fitness advantage and cluster composition, i.e., the proportion of cell types, determines the fitness in environmental conditions where a division of labour or metabolic interactions provide an advantage8. Understanding the biological knobs that prescribe the steady-state size and composition of these clusters is imperative to gain insights into the evolution of multicellularity.

Growth of differentiating multicellular clusters involves three processes i) Proliferation by addition of cells, ii) Differentiation by phenotypic switching2 iii) Fragmentation due to prevailing physical forces9. Previous work has highlighted genotypes that control maximum cluster sizes and internal physical forces. For example, changes in cell shape4 or strength of cell-cell connection allow clusters to accommodate, on average, a more significant number of cells per cluster. While it is well understood what the principal targets of natural selection for multicellularity in yeast are, a comprehensive understanding of the principles that determine the steady-state dynamics of growth and fragmentation in these clusters is lacking. Some of the frequently targeted modules in these evolution experiments increase cell-to-cell connections' strength, i.e., ACE2, CTS1, GIN4 or are involved in cell shape regulation10, i.e., CLB2, ARP5. How alterations in these cellular features affect the steady-state growth dynamics of these clusters remains largely unknown. Moreover, it is important to appreciate that the persistence and prevalence of any cellular interactions in more sophisticated multicellular systems like metabolic cross-feeding or division of labour would depend on the dynamics of growth and division than the cluster sizes alone. A predictive model of the multicellular growth would allow exploration of parameter space that give rise to specific cluster size and composition.

**A dynamic-network approach to model the growth of multicellular clusters**

We use a dynamic network model, MultiCellNet representing clusters' growth dynamics with cells as nodes and their connections as edges. We program the dynamic network to grow proportional to cell division rate () and fragment at a specific edge when a certain set of conditions are met. Using this approach, we explore multiple models with different rules of cluster fragmentation. Crowding around a specific cell in multicellular clusters leads to the development of steric forces due to geometric constraints. We capture such a constraint in the parameter 'kissing number ()' derived from geometric limits on the organization of spheres11. Following the development of steric forces, one or more cell-cell connections fractures to release the strain. The fracture can either happen probabilistically, i.e., each connection has a probability of breaking based on an underlying probability distribution or deterministic, for example, the oldest connection breaks. The fracture partitions the cells in the parental cluster in some ratio determined by the modality of fracture (**Figure 1A**). Such a cyclic process occurs asynchronously in several clusters, each at a different stage of growth in a population. At a steady state and in a well-mixed system, this gives rise to a distribution of cluster sizes. To test whether individual clusters encode the complete information needed to give rise to the entire distribution, we used Fluorescent Activated Cell Sorting (FACS) to separate all four quartiles (Q1, Q2, Q3 and Q4) of the distribution obtained from the *cts1∆* strain that forms clusters of size ranging from 3-10 cells8. We then let the four populations grow and independently reach a steady state (Figure S1A). All four populations individually recapitulate the entire distribution, strongly suggesting that it is generated from an underlying process with deterministic characteristics.

To understand the exact process that emits the entire distribution, we consider five possible modalities of fracture (**Figure 1B**). There are two broad possibilities on what limits the size of a cluster: i) The clusters grow and reach a packing limit beyond which steric forces onset, and consequently, it fractures to relieve the strain. We call this modality 'Steric Limited (S)' ii) The links between cells have a half-life in the order of cell doubling time, and therefore within a few divisions of growth, the links fractures due to chemical degradation. This process does not involve the onset of physical forces, and we call this 'Non-Steric Limited (NS)'. Each cluster has several cell-cell links, which is non-trivial as to which one most likely fracture. The choice can be purely deterministic such that in the S modality, the oldest link breaks as the cells it links have reached the maximum number of cells they can be surrounded with. We call this model SO to describe a deterministic fracture of the oldest link in a steric-limited fashion. We did not consider a non-steric limited version of the oldest link fracture as it would only produce single cells.

Another possibility is that an underlying stochastic process assigns a fracture probability to each link. Assuming all the links are equally vulnerable to fracture, the probabilities can be derived from a uniform distribution. In this case, once the cluster reaches the limit of breaking either due to steric forces (S) or due to degradation of linkages (NS), one or more randomly chosen link fractures. We call these models SR and NSR. In an alternative setting, the chronological age (A) of the link matters and therefore longer it has been present in the cluster, the higher the probability it fractures. We consider an exponential distribution to pre-assign the duration for which a link can remain intact before the onset of steric forces. Such a distribution is parameterized by a single rate parameter (δ). We call this model SA.

All the models encompass three significant parameters, i.e., cell division rate (), kissing number () and link-breaking rate (δ). Now, all five models involve at least 2 of 3 input parameters. To test which of the models best explains the observed cluster size distribution in experiments, we take two approaches i) We engineer strains where we can control biological equivalents of model parameters ii) Develop an imaging pipeline that allows counting the number of cells per clusters across a population of clusters previously difficult due to light scattering in intact three-dimensional clusters. We screen (**Figure S2**) and obtain a set of genes whose expression levels alter one of three parameters independently and in an inducible manner. 8For example, controlling the expression of BUD412 protein (**Figure 1C**) in a strain deleted for ACE2 switches the budding pattern from axial (ON) to bi-polar (OFF). This allows for switching the kissing number limit between a lower and upper bound. We avoid controlling the expression of genes involved in cell-shape regulation to alter kissing numbers due to the non-linear relation between protein levels and kissing numbers (Figure S2).

Similarly, tuning the expression of chitinase CTS1 changes the hydrolysis rate of chitin and allows for alteration of the link-breaking rate (δ). To gain control of the cell division rate of cells in the cluster (), we alter the concentration of the carbon source in a carbon-limited minimal media13. Finally, we compare the simulated cluster size distribution obtained from all the models with the experimental distribution (**Figure 1D**) to assess the performance of the models.

We find that a steric-limited and age-dependent fracture model (SA) captures the experimentally observed distribution. We find that the cluster size distribution follows Erlang's distribution parameterized by all three parameters in the model. Furthermore, we find that cluster size distribution follows a rheostat-like model where cell division rate sets a fundamental threshold for link breaking rate such that their relative values determine the mean size of clusters. In this model, the kissing number independently controls the upper limit on cluster size. We apply the modelling framework to understand how a differentiating cluster with two cell types maintains compositional homeostasis and whether the composition can be experimentally tunable. We find that by adjusting switching rate between the cell types and their relative growth, the composition of the clusters can be controlled. Moreover, the homogeneity or precision in the composition across different clusters in a population is increased by increasing the kissing number. Overall, we find a growth model for multicellular clusters that would help us understand the evolvability of multicellular phenotypes, such as size and composition and enable rational engineering of multicellular clusters for synthetic biology applications.

**Results**

**High-throughput pipeline enables quantification of cell number per cluster across conditions and genotypes**

Simulations using the proposed model set produce a cluster size distribution in number of cells per cluster. To accurately identify the agreement between the models and the experimental distribution, it is necessary to quantify the number of cells per cluster across many clusters. It is challenging to image individual cells in an entire three-dimensional cluster due to excessive scattering and interference by the yeast cell wall. Moreover, light-sheet microscopy might help overcome this problem but would fall short in throughput15. We need enough clusters to perform statistics on them to compare model-derived and experimental distribution. The clusters were stained with calcofluor white (CW) to overcome this challenge and squeezed between a coverslip and a slide. The fluid volume between the coverslip and slide was optimized so that the cells formed a monolayer without lysing (**Figure 2**). Furthermore, we performed a timelapse of clusters disintegrating to a monolayer to check for any lysis event or mixing of cells from different clusters **(Video S1**). By optimizing the density of clusters in a field of view, we spatially resolved cells belonging to a specific cluster in a monolayer (**Figure 2B and Figure S3**). We used RCNN-Mask trained on retina images of previously top-segment yeast cells using images obtained from CW channels (**Figure 2B**). We collected multiple fields of view (>60 FOVs, i.e., ~200 clusters) to obtain a distribution accurately. The minimum number of clusters to be obtained to claim a detectable difference in two conditions or genetic backgrounds was calculated based on Cohen's d (**Figure S2A**). For all the experiments, the expected effect size was <0.4) and therefore, 200 clusters were enough to derive a representative distribution. The centroid of cells in segmented images was estimated and used to cluster cells based on their spatial location using DBSCAN (Density-Based Spatial Clustering of Applications with Noise). The number of cells corresponding to each cluster was used to generate a discrete distribution of cluster sizes (**Figure 2**). We obtained reproducible cluster distributions for various genetic backgrounds (corresponding to size distributions) (**Figure S2B-D).**

**A steric-limited age-dependent fracture (SA) based dynamic network model predicts cluster size distributions**

To test model predictions, we used a one-factor-at-a-time approach. For testing the effect of kissing number () on cluster size distribution as obtained from both experiments and simulations, we varied the levels of Bud4p protein in a strain with *ace2∆* background. For this, we used a beta-estradiol-based inducible system16, which allows tight regulation while minimizing the effect of cell growth by replacing the native BUD4 promoter. High expression of BUD4 leads to axial budding (as observed in wild-type background), while shutting the expression off switches the budding pattern to bipolar. A bipolar pattern allows yeast two consecutively bud from distal poles, allowing the cells to have a higher kissing number. Using the pipeline described earlier (**Figure 2**), the cluster size distributions corresponding to BUD4-ON and BUD4-OFF were obtained. The maximum number of cells in a cluster obtained in the case of BUD4-OFF was between 128 and 256. This implies an expected kissing number of 8, as exceeding this would produce clusters with number of cells greater than 256. Similarly, for BUD4-ON clusters, the maximum number of cells per cluster was between 16 and 32, implying a kissing number of 5. The doubling time/cell division rate () of the clusters was estimated from the growth curve (**Figure S7B**). The link breaking rate () for the SA model was set to a very high value ( such that link breaking does not fundamentally limit the size of the clusters. The cutoff value for was estimated in the following manner:

Formation of cluster close to the maximum size (set by ) takes time where is the cell doubling time assuming all the cells in a cluster divide at the same rate and is the number of generations lapsed. In SA model, the time for which a link remains intact is derived from an exponential distribution, . Now is the probability that the oldest link fractures at the instant the cluster reaches the kissing number and is the probability that the cluster has reached a size close to the kissing number limit. For SA model, where the cluster size is solely limited by the kissing number, should tend to 1.

To summarize, this means the probability,, such that the link fractures only when the cluster reaches the limit of breaking is 1. If this is not true, then either the cluster breaks before the cluster reaches the steric limit, which violates the basis of the SA model. This implies that:

As there is no simple analytical solution for this, expanding by Taylor's approximation and for the unique value of ,

Solving the quadratic for a unique value of ,

Where .

Therefore, for values above, the model remains steric-limited. This result is significant because there is not any direct approach to measure the experimental link-breaking rate in a reliable way for any of the genotypes described earlier. Setting the link breaking rate above the threshold ensures the definition of the SA model remains consistent, and the breaking rate does not appear as a free parameter in the model.

Based on the values of , the models were parameterized. On overlaying discrete distributions obtained from experiments from two conditions, i.e., BUD4-OFF and BUD4-ON with Kernel Density Fits from cluster sizes obtained from simulations, the SA model shows a strong agreement in predicting the cluster size distributions for uninduced Bud4p levels (**Figure 3B and 3C**) and induced Bud4p levels (**Figure 3E and 3F**). Increasing the kissing number by shutting down BUD4 expression also strongly suggests that the cluster size is steric-limited. We image the budding pattern using confocal laser scanning microscopy with CW-stained clusters (**Figures 3A and 3D**). Budding close to the previous bud-scar, i.e., axial budding, lead to the early development of steric forces. Therefore, clusters collapse at a smaller cluster size compared to bipolar budding. Model SR produces a cluster-size distribution that is larger than experimentally realized. SR model assigns an equal probability of breaking to all the edges. The breaking of the oldest edge produces clusters of equal sizes, whereas the breaking of the newest link produces clusters of highly asymmetric clusters. The theoretical number of cells per cluster if the newest edge breaks for a kissing number of 8 would be 254 and 2. A size of 254 is not realized in the model. Similarly, for a kissing number of 5, the resultant clusters would have sizes 30 and 2. Therefore, model SR fails to capture the experimentally observed distribution.

Other competing models fail to show any correspondence with the observed experimental distribution. For example, Model NSA produces very large sizes of clusters (>500 cells per cluster) that are not realized experimentally in any conditions. Model NSR produces cells with only one cell as a random choice to break one of the edges starting from clusters having just two cells would produce singlets. Model SO produces a bimodal distribution of cluster sizes corresponding to 128 and 64 cells per cluster. Breaking at the oldest node will always produce clusters of equal sizes as the cluster is symmetric about the oldest edge. Therefore, the clusters would either have 128 cells per cluster when they are about to break or 64 cells per cluster after they break.

To further test whether the assumption that the probabilities of the link breaking at a specific time is derived from an exponential distribution, we took the following experimental approach. We used a strain where the expression of Cts1p encodes for the significant chitin hydrolase is tunable. Titrating the expression of this protein using beta-estradiol in the dynamic range allows for changing rate at which link breaks due to chemical hydrolysis, which alters the statistics of the underlying distribution. If the rate of hydrolysis of the link (realized as in the simulation) determines its probability of fracture where is fixed, then increasing Cts1p expression should lead to the collapse of the links at a higher rate. As the cluster breaks after divisions, the cluster size distributions follow the *Gamma* distribution if the breaking rate ( is the exponential distribution. Gamma distribution cannot be directly used as it includes zero, which is undefined for a distribution having size as a random variable. The appropriate distribution is a zero-corrected version of the *Gamma* distribution, Erlang's distribution. For random variables that follow Erlang's distribution, the mean and standard deviation are linearly correlated with the slope . We induced the expression of Cts1p in a graded manner and generated cluster size distribution for each induction level using the method described previously (**Figure 3G**). Consistent with the hypothesis, the means and the standard deviations follow Erlang's distribution (**Figure 3H**). Each point represents a given ß-estradiol concentration that is used to titrate Cts1p levels. There could be two mechanistic reasons why exponential distribution captures the breaking probabilities of the link i) The oldest edge has a higher probability of breaking than newer ones due to chemical degradation over time ii) The oldest edge also branches off to the greatest number of cells, and therefore most steric forces would act on the oldest link9. In either of the cases, an age-dependent septum fracture model would explain the probabilistic breaking. It could also be a combination of both. Given that both influence the link-breaking probability similarly, an exponential distribution would still capture the effect of both. Decoupling the effect would require careful measurement of the tensile strength of all the links in question.

In previous experiments, we can only claim the resultant distribution obtained from the model, and experiments show strong agreement. It is very possible that clusters in the model follow an entirely different breaking mechanism while producing the same distribution of cluster sizes. To further test if this is the case, we performed time-lapse imaging of clusters growing and fragmenting. We simultaneously simulate the dynamic network model and longitudinally track the cluster sizes in a lineage. We define 'mother' as the cluster from which two or more 'daughter' clusters break off. The ratio between the size of the more minor of two daughters (Sdaughter) and the mother cluster (Smother) gives a rough indication of whether the breaking mechanism is biased towards the oldest link. If the ratio is close to 0.5, the cluster breaks off from the oldest link. The hypothesis suggests that the distribution will be biased towards higher ratios. To perform this, we used a cluster with Bud4p expression shut off and constitutive expression of CFP. The clusters were imaged at a low magnification to increase the depth of field and allow the collection of a maximum number of photons (**Figure 4A and Figure 4B)**. Here, we assume the net fluorescence from a cluster as a proxy for the number of cells per cluster. Consistent with the hypothesis, we find that the ratio distribution is loaded towards higher ratios (**Figure 4C**). We find a strong agreement when we perform similar estimations using the simulated dynamic network model (**Figure 4D**). This establishes that the SA model mechanistically follows the experimentally realized breaking mode.

**Cluster size homeostasis is determined by a rheostat-like model where the growth rate caps the maximum breaking rate**

We hypothesized that if the rate of addition of cells to a growing cluster exceeds the rate of breaking determined by kissing number and link breaking rate, the size distribution will shift towards bigger clusters (**Figure 5A**). An underlying assumption is that the faster addition of cells only reduces the time taken for a cluster to reach close to the kissing number () while the link-breaking rate () remains unaltered. The link-breaking rate () is an intrinsic property the mother-daughter septum determined by the location of the link within the cluster and its mechanical properties. Intuitively, this suggests that if the rate at which cluster fragments, i.e. determined by both link breaking rate and kissing number, exceeds the cell division rate of the constituent cells, on average, the size of the clusters will be shrunk. Therefore, values for and are constrained to a specific parameter space for a population of clusters to attain a specific mean size. The growth of clusters at low glucose concentrations (0.025% or 0.05%) reduces the mean cell division rate due to nutrient limitations (**Figure S7B-D).** As the principal transcription factor17,18 for chitin hydrolysis ACE2 is deleted in the strains, we have decoupled the effect of slow growth on regulated chitin hydrolysis. Therefore, at different concentrations, the primary effect on the cluster size distribution is through the nutrient limitation induced slow down of growth. We measured the cluster size distribution for the strain with Bud4 expression shutoff in an *ace2∆* background and growing exponentially in different concentrations of glucose. Consistent with the hypothesis, the cluster size distribution is shifted to bigger clusters (reflected in the mean of the distribution) for growth on higher glucose concentrations (2% ad 0.2%) compared to lower concentrations (**Figure 5B**). To further explore the causality between growth rate and the mean cluster size, we leveraged the model to explore the parameter space and ask how and influence the relation between and mean cluster size. Simulations suggested that, for a fixed cell division rate, the mean cells per cluster increase with link breaking rate in a sigmoidal fashion (**Figure 5C**). Interestingly, the half-maximum vean number of cells per cluster is proportional to the doubling time of cells in the cluster (**Figure 5C)**. The maximum value (S bar max) is set by the kissing number of the cluster in question. Based on this information, we propose a rheostat-kind of model where the relative values of the link breaking rate () and cell division rate ( alter the mean size of the clusters. In contrast, the kissing number () sets the maximum value. We used a strain in which CTS1, which encodes the primary chitinase in yeast, is under the control of a beta-estradiol inducer to test the validity of this model. In this strain, we can regulate the rate of chitin hydrolysis (inversely proportional to ) and grow the clusters in varied glucose concentrations to alter the cell division rate (). Using this system, we experimentally explored the parameter space of and . We used ß-estradiol inducer levels as a proxy for link breaking rate (), i.e., lower concentrations of ß-estradiol lead to increased . The cell division rate () was calculated from the growth curves for bulk cultures of the clusters in different glucose concentrations in minimal media (**Figure S7C-D**). For this specific experiment, we considered all the combinations of three different ß-estradiol concentrations, i.e., 0, 4, and 8 nM, with three different cell division rates (), i.e., 0.22 hr-1, 0.18 hr-1, 0.12 hr-1. The number of cells per cluster was estimated using the method described previously (**Figure 2**)**.** The mean number of cells per cluster was used as a proxy for the effect of cell division rates and link-breaking rate on cluster size distribution. Consistent with the proposed model, the cell division rate sets a threshold for the link-breaking rate to alter the size of the clusters (**Figures 5D and 5E**). The effect of concentration on cluster size distribution is independent of the quality of the carbon source, as galactose produces a similar effect **(Figure S7D**). At higher cell division rates, i.e., 0.22 hr-1, lower concentrations of ß-estradiol, i.e., 0 nM (corresponding to high ), lead to increased cluster size than high concentrations (corresponding to lower ). This effect is diminished at lower growth rates, i.e., 0.12 hr-1 (lower threshold), consistent with the rheostat-like model. The rheostat model links all three parameters in the dynamic network model in an empirical form. It also highlights the fundamental principles that allow a physically connected network of cells to reach a given size distribution and how the underlying parameters interact. We used the dynamic network based on the steric and age-dependent (SA) model to explore behaviourvior of multicellular clusters in which cells switch from one phenotypic state to another and ask how they maintain compositional homeostasis during growth-fragmentation cycles.

**Higher kissing number minimizes compositional heterogeneity in differentiating multicellular clusters**

When multicellular growth occurs with irreversible differentiation from germline to somatic state, proportionate inheritance of these two cell states is not guaranteed. This leads to the disproportionate distribution of two cell states amongst all multicellular clusters. Moreover, clusters ending with only somatic cells cannot reproduce a two-state cluster. In environmental setups where the division of labor is necessitated, such heterogenous clusters will penalize fitness over time if constantly produced. Therefore, mechanisms that minimize the formation of disproportionate or sterile clusters would provide a fitness advantage to the population. Previously, our research group reported a system where the somatic cells express invertase (SUC2), whereas the germline harbours a resistant marker that allows survival on cycloheximide8. Invertase produced from somatic cells breaks down sucrose and allows the germline to survive, propagating genetic information. Cre-Lox recombination allows the excision of a genetic segment that harbors the resistance allele and the recombination of a promoter and *suc2* gene. Low-level induction of Cre recombinase leads to stochastic and irreversible differentiation of cells germline state to somatic state. A fluorescent marker (YFP for somatic and RFP for germline) is linked to either state to allow quantification of the two cell types in a cluster. We define the switching rate from a germline-like state to a somatic state as And the relative growth advantage of somatic with respect to germline as . The parameter can be experimentally modulated by changing the induction of Cre-recombinase by ß-estradiol and can be controlled by changing the cycloheximide concentration.

In the context of a multicellular cluster, if the differentiation event happens in the same branch that has the highest probability of breaking, a sterile cluster will be formed (**Figure 6A**). If the number of cells connected to a given cell is low i.e., kissing number () is low, the probability of the differentiating branch and the branch that will fracture coinciding is high assuming these two processed to be independent. In the case of a high kissing number, the corresponding probability will be reduced due to lower chances of both events coinciding. We modified the MultiCellNet model with steric and age-dependent modality of fracture to incorporate the dynamics of differentiation. Briefly, when a new node is added to a growing network, is assigned a phenotypic state (germline or somatic) based on the value of . Moreover, somatic state cannot switch back to germline and therefore . At a value of i.e., no relative growth advantage of germline, increasing kissing number only marginally reduces the fraction of clusters that are sterile (**Figure 6B**). On increasing the relative growth advantage of germline-state, at , the effect of kissing number on fraction of sterile clusters is more pronounced (**Figure 6B**). Simulations also predict that increased kissing number also increases the fraction of clusters that have both the cell types (mix).

To test the hypotheses experimentally, we engineered the differentiating clusters by additionally knocking out either BUD219 (buds at random sites on cell surface) or BUD320 (buds at bipolar sites) in a *cts1∆* background. These clusters have increased kissing number reflected in the maximum attainable size in the cluster due to switch from axial budding to bipolar or random budding. The fraction of clusters that have exclusively germline or somatic or a mixture of both can be identified by flow cytometry based on total RFP or YFP signal emitted for each cluster (**Figure S8**). We measured the fraction of each cluster type, i.e., germline, somatic, and mix for increasing concentrations of cycloheximide for both *cts1∆*, *bud2∆ cts1∆ and bud3∆ cts1∆* clusters. Here, glucose was used as the sole carbon source to eliminate the influence of sucrose in distorting the distribution due to selection effects7. Increasing the kissing number, clearly reduces the fraction of sterile clusters, when compared between *cts1∆* and *bud2∆ cts1∆* clusters with equal relative growth advantage (**Figure 6C**). It also leads to a marginal but significant increase in fraction of clusters that are mixed. The alteration of sterile and mixed fraction for all the conditions noted is clearly visible with light microscopy (**Figure 6D**). One potential disagreement between the model simulations and the experimental results is reflected in the fraction of clusters that are exclusively germline. While the model predicts that at steady state that this fraction reduces at high kissing numbers, experiment suggest that there is an increase. This either might be an unprecedented effect on either the drugs used or the genetic background. Moreover, the perturbations might be altering the absolute cell division rates or germline or somatic state even while maintaining a constant

Furthermore, we were interested in testing how do the model parameters influence the ratio of both cell types in individual clusters, the variability in this ratio across clusters in the population and whether it can be modulated in a predictable manner (**Figure 7A**). In settings where metabolic division of labor or cross-feeding is prevalent, ratio of two cell-types in a cluster becomes particularly relevant. An imbalanced ratio would underfeed one of the two cell-types and would compromise the fitness of the whole system.

We modified the pipeline used to measure number of cells per cluster (**Figure 2**) to enable estimating the number of cells in a cluster that belong to a specific cell-type i.e., germline or somatic. This allowed us to measure the fraction of a given-cell type in individual clusters based on YFP or RFP signal across hundreds of clusters in a population and for several different perturbations. First, we tested how parameters specific to differentiating clusters i.e., and alter this mean fraction of somatic cells per cluster. We club the net effect of these two parameters in = . Here, the term captures the effect rate of formation of somatic cells and for the rate of growth. Therefore, describes the net effect of formation and survival of the somatic cell-type. MultiCellNet predicts that the mean fraction of somatic cell-type in clusters monotonically increases with the net switching rate (**Figure 7B**). Any desired ratio can be attained by fixing the system at a given net switching rate () as per model predictions.This relation seems to be largely independent of both the kissing number () and link breaking rate (). To test the validity of this relation, we set-up an experimental system to alter the net switching rate by changing its constituent terms through ß-estradiol and cycloheximide treatment. We used three different switching rates () corresponding to 3 nM, 10 nM, and 30 nM ß-estradiol and three different relative growth advantages () corresponding to 0 nM, 150 nM, and 300 nM cycloheximide. We measured the fraction of somatic cells for hundreds of clusters per sample, for all combination of strains, switching rate and relative growth advantage i.e., total of 9 conditions. Consistent with model predictions, the somatic fraction scaled monotonically with net switching rate independent of the strain background (**Figure 7C**). The fraction of somatic cells changed from 0.3 for lowest to 0.55 for highest . Given that we collected fraction of somatic cells for hundreds of clusters, we explored the cluster-to-cluster variability for various strain backgrounds. As fraction is bounded between 0 and 1, we measured Gini coefficient to measure cluster-to-cluster variability in fraction of somatic state. Gini coefficient is adopted from economics used to quantify income variability in nations21. We found a strong negative correlation between mean somatic fraction (**Figure 7D**) and the Gini coefficient for all the conditions measured (r2=0.86). This implies a scaling relation between mean and variability like what is observed for gene expression mean and noise. A higher somatic fraction automatically leads to a reduction in cluster-to-cluster variability. A similar quantification was also made using conventional measures like coefficient of variation, which is used for unbounded measurements, and we obtained a similar relation (**Figure S9B**). We quantified the distribution of Gini coefficients for all the three strain backgrounds to check for any differences. Strikingly, *bud3∆ cts1∆* clusters had significantly lower Gini coefficients compared to both *cts1∆* and *bud2∆ cts1∆* clusters (**Figure 7E**). It is difficult to purely assign this effect to kissing number as *bud2∆ cts1∆* clusters showed a higher Gini coefficient distribution compared *cts1∆* distribution. A plausible cause for this effect could be the budding pattern associated with *bud3∆ cts1∆* i.e., it buds bipolarly. Clusters with *cts1∆* and *bud2∆ cts1∆* bud axially and randomly respectively. Both the modalities of budding lead to a higher coincidence of 2nd bud emergence close to the previous one. In axial budding (*cts1∆*), it is almost certain for such an event to occur. In random budding as location of 1st and 2nd bud are independently chosen, the chances that two consecutive buds coincide is in no way minimized. Contrastingly, only in the case of bipolar budding, two consecutive buds are formed in distant pole and hence reducing the probability of coincidence event. This observation hints at the importance of geometric position of buds in determining the variability of cell-state fraction which is not captured in our dynamic network model. This highlights a potential limit of the current modelling approach and calls for incorporation of geometric position for determining phenotypes like variability.

**Discussion**

Evolved yeast clusters throw light on at least a couple of scenarios that would have selected multicellularity1. Such model systems help inquire what molecular features control cellular organization in the multicellular system as seen in more complex organisms across the tree of life. The ease of genetic manipulation makes yeast an excellent tool for understanding the dynamics of how single cells form multicellular structures. Complex multicellular organisms have evolved mechanisms to control the proportion of their cell types to provide them with reproductive advantage. Current research argues how mechanisms that control cell cycle combined with morphogen sensing and lateral inhibition22 allow them to maintain robust ratios despite stochasticity in cell fate decisions. The absence of correct cell-type ratios can compromise fitness in scenarios where metabolic interactions prevail23. For example, in the case of obligate resource cross-feeding, two-cell types must be in the same cluster. How do cells achieve such robustness despite the stochastic nature of the underlying processes is an open question?

Here, in a simple eukaryotic, multicellular system, we show that the modality of cluster fragmentation can regulate the size distribution in clusters using a dynamic network modelling approach. The model also does not take any parameters related to the exact mechanics of fracture or geometry of cell size and position and recapitulates the experimental cluster size distributions across a wide array of conditions. While we experimentally manipulate the budding pattern or chitin hydrolysis rate using approaches idiosyncratic to *Saccharomyces cerevisiae*, the model(s) parameters are general, i.e., kissing number or link breaking rate and are thus applicable to any multicellular network of cells. Previous work has also shown that the organization of cells in clusters follow information entropy maximization principles irrespective of the organism or cell type15. We, therefore, hypothesize that such a model can capture growth and fragmentation dynamics in other multicellular systems after adjusting the parameters to specific cases. Applying the models to differentiating organoids/spheroids or organoid tumor models would be interesting to see if growth and fragmentation can be modelled by a dynamic network model like the one proposed in this study.

Moreover, this study also highlights how a single perturbation i.e., kissing number can concomitantly i) increase the size of the clusters ii) reduce the fraction of sterile clusters iii) reduces the cluster-to-cluster variability across a set of conditions. We speculate that these perturbations would have provided multiple advantages to organisms at once allowing a hard sweep of alleles functionally analogous to *bud3∆* or *bud4∆* during natural selection. Moreover, our observations from Figure 5, highlight that faster growth of individual cells lead to a shift in the cluster size distributions towards bigger clusters. We also speculate that such a relation would have allowed formation of bigger clusters without any direct cost to cell division rate which potentially supports the plausibility of multicellularity as a trait.

A limitation of this study is the assumption of tree-like growth of the cluster of cells which is idiosyncratic to multicellular yeast. In more complex life forms, the cells in the clusters are cemented together by an extracellular matrix comprising collagen or other binders24. This allows cells to form connections with other surrounding cells, unlike yeast clusters where a cell is only linked to the mother cell. Moreover, cells in complex multicellular organisms can sense mechanical forces25 around them and control cell division. This allows them to form 3D structures like organs using morphogenetic clues and having feedback control on their cell cycle. Given the ease of adding more features to the dynamic network model i.e., as attributes to nodes and edges, it would be interesting to further explore and see what other features need to incorporate to explain multicellular phenotypes in more sophisticated form of development and organization. For example, would incorporating information about lateral inhibition26 as a node attribute explain how mechanosensitive circuits modulate cluster size when a high level of precise in required? Or would incorporating morphogen sensing27 as a node attribute based on geometric location allow prediction of patterning?

**STAR Methods**

**Key Resource Table**

|  |  |  |
| --- | --- | --- |
| Reagent or Resource | Source | Identifier |
| **Chemicals, Peptides, and Recombinant Proteins** | | |
| Beta-estradiol |  |  |
| Concanavalin |  |  |
| Cycloheximide |  |  |
| Yeast Nitrogen Base |  |  |
| Complete Supplement Media |  |  |
| Dextrose |  |  |
| Galactose |  |  |
| Ethanol |  |  |
| Calcofluor White |  |  |
| **Deposited Data** | | |
| GitHub |  |  |
| Zenodo |  |  |
| **Experimental Model: Strains or Organism** | | |
| Yeast Strains | Table S1 |  |
| **Recombinant DNA** | | |
| ERBD plasmid | Addgene |  |
| Beta-estradiol plasmid | Addgene |  |
| Plasmids | Table S2 |  |
| Oligonucleotides | Table S3 |  |
| **Software and Algorithms** | | |
| DeepRetina (RCNN) | GitHub |  |
| Anaconda |  |  |
| MATLAB |  |  |
| SLURM |  |  |
| Cytoscape |  |  |
| **Others** | | |
| 384 glass bottom plates |  |  |
| CellASIC ONIX2 system |  |  |
| Glass slides |  |  |
| Coverslip |  |  |

**Resource Availability**

Further information or request for resources should be forwarded to the lead contact, Andrew Murray (awm@mcb.harvard.edu)

**Materials Availability**

Yeast strains and plasmid built for this project are available on request. Please contact the lead contact for further details.

**Data Availability**

All scripts used for processing data or generating figures are hosted on GitHub. All raw data from microscopy experiments have been submitted in Zenodo. Any inquiry must be directed to the lead contact.

**Experimental Model and Subject Details**

*Yeast strains and media*

All experiments were performed using strains constructed from modified W303 background28. Briefly, the functional version of BUD4 allele from S288C was engineered in the endogenous locus in W303. Beta-estradiol transcription factor was introduced at HIS3 locus through an integrative plasmid16. The integration was confirmed by performing PCR using primers specific to the locus and the transcription factor. Genes of interest were put under the control of beta-estradiol promoter by replacing the endogenous promoter by homologous recombination using 40 bp homology29. The replacement was confirmed by performing PCR with primers specific to the locus and the promoter construct. Constitutively expressed fluorescent markers were introduced at HO locus using homologous recombination through KanMX marker and 40 bp homology30.   
  
Gene deletions were performed by amplifying KanMX resistance marker with 500 bp upstream and downstream sequence from yeast deletion collection31,32 and introduced in the background of interest by homologous recombination. Confirmation PCR was performed to both check the integration of the resistance cassette and removal of target gene. Replica plating was performed for each successive transformation to confirm retention of previous markers.

Strains were grown in synthetic minimal media supplemented with carbon source at desired concentrations. Media were prepared freshly by diluting 10X stocks of the refrigerated constituents and prewarmed before inoculation. For every experiment, strains were freshly streaked out from -700C glycerol stocks (15%) and allowed to recover for at least 2 days at 300C. Strains were cultured in desired media for at least 10 generations before any measurements were performed. For specific experiments, strains were grown longer than 10 generations by maintain them in exponential phase for size distribution to reach steady state. All the experiments were performed with two biological replicates for high-throughput imaging of size distributions. Every experiment was repeated at least thrice on independent days and replicate correlations have been described in supplementary figures.

**Methods details**

*Estimating number of cells per clusters*

Strains were grown in indicated media for at least 10 generations to reach steady state distribution. For performing imaging, exponential grown cells were incubated with 1 volume of calcofluor white without 10% Potassium hydroxide for 1 minute at RT. Approximately, 2.8 uL of the mixture was squeezed between a microscope slide and coverslip of dimensions 20 mm X 20 mm. Coverslip and microscope slide were cleaned with air jet before experiments to remove dust particles. Nikon Eclipse TiE inverted fluorescence microscope with Hamamatsu EMCCD camera was used for performing all imaging operations. The microscope was controlled through MetaMorph software. An automated script written in MetMorph to scan 49 positions on the slides in an arrayed fashion. The density of cultures was optimized to yield at least 4-5 clusters per field of view. Monolayer formation was manually verified for each sample by imaging z-stacks and confirming absence of more than one layer of cells.

*Timelapse imaging*

All timelapse imaging were performed in the microscope set-up mentioned earlier. Concanavalin-A coated wells in a 384 well plate was used to perform lineage tracing of cluster growth and division. 10X objective magnification with 10 mm working distance was used in this set-up. A constitutively expressed CFP was used as a proxy for total number of cells per cluster. The depth of field was calculated using excitation wavelength and magnification values. The plates were filled with at least 40 uL of media to reduce evaporative loss. Strains were grown in specified media for 10 generations and were spun down on a 384 well plate at 300 g for 2 mins. Images were taken every 10 mins for a 12 h period.

*Confocal Laser Scanning Microscopy of 3D clusters*

Zeiss LSM900 was used for performing 3D imaging of multicellular clusters. Briefly, cells were stained with calcofluor white and loaded onto a concavity slide in 0.2% agarose to keep clusters stationary during imaging. The clusters were imaged using DAPI channel with a z-stack width of 0.2 um. The 3D images were rendered using Fiji (ImageJ) and custom scripts.

*Flow cytometry and sorting*

All flow cytometry experiments were performed in FACS LSR II. The photomultiplier tube (PMT) voltages for forward and side scatter were optimized for each experiment to record maximum signal from the biggest clusters. The area signal from forward scatter (FSC-A) was used as a proxy for cluster size for preliminary measurements. The flow cytometry tubes were vigorously vortexed to obtain a uniform suspension of clusters before loading the samples.

FACS was performed on multicellular clusters using FACS Aria II using signal from FSC-A channel. The FSC-A distribution was segmented into 4 quartiles and sorted into 4 different collection tubes at RT with prewarmed minimal media. The cluster fractions were grown in flasks at 30 deg C for 48 hours. The size distribution evolution over time was recorded every 12 hours using flow cytometry.

*ß-estradiol titration*

Strains were grown overnight in minimal media without beta-estradiol. They were sub-cultured next day into minimal media with desired beta-estradiol concentrations. Beta-estradiol stocks were prepared in 100% ethanol and stored at -200C. The stock was diluted to desired concentrations in minimal media. All measurements were performed after overnight growth (at least 10 generations) in presence of inducer. The dynamic range of inducer concentrations was estimated by performing a log-scale dose-response curve with the output as cluster size distribution.

*ß-estradiol Cre induction and differentiation*

Strains were freshly streaked out from -700C glycerol stock. Single colonies were picked and checked for spontaneous Cre recombination events. Colonies with pure germline state were grown in complete synthetic media (prepared from 10X stocks of constituents) overnight and diluted 100-fold into fresh media supplemented with desired concentrations of beta-estradiol (inducing Cre recombinase expression) and cycloheximide. Cycloheximide stocks were made in 100% ethanol and stored at -200C. The strains were grown for at least 14 generations (two transfers) to allow clusters to reach steady state fraction and size. Differentiation was confirmed by microscopy and flow cytometry.

Ratio of cell types in individual clusters were determined using a modified version of method described in Figure 2. Instead of calcofluor white, a constitutively expressed CFP present in both germline and somatic was used as segmentation marker. Number of cells of a given cell-type was estimated by pipeline described in the quantification and analysis section.

**Quantification and statistical analysis**

*Dynamic Network Simulations and Analysis*

Custom scripts were written in Python to perform dynamic network simulations using NetworkX library. Briefly, a recursive program was run in which nodes (cells) were added every division time to extant nodes. Cell properties were encoded in node attributes and link properties were encoded in edge attributes. A function was used to update the attribute based on a trigger event: i) whether a kissing number was reached ii) the links reached a point of breaking. The network growth is initiated from a single cell but continues till 10 generations. Disconnected modules in the network were determined by finding set of nodes which don't have connections with outside nodes. A disconnected module was identified as a cluster and number of cells was determined. Every simulation was performed for multiple random number seeds. Same random number seed was used for comparing effects of two different parameter sets on network properties.

*Segmentation and estimation of cluster attributes*

Segmentation of cells in clusters were performed using a yeast optimized pipeline based off DeepRetina. The pipeline was executed in supercomputing cluster (FASRC) using custom SLURM scripts and images acquired were directly transferred for segmentation without any preprocessing or selection. The aspect ratio of images was preserved, and images acquired through DAPI channel or CFP was used for segmentation.

MATLAB 2022B was used for image processing and analysis. Briefly, centroids corresponding to cells in a cluster was determined and used for performing DBSCAN (Density Based Spatial Clustering And Noise). DBSCAN assigns the cells to specific cluster based on its geometric location. Clusters with cells close to boundary (in the 50-pixel proximity to boundary) were rejected from downstream analysis pipeline.

For estimating fraction of somatic cells per cluster, signal from YFP channel was used. Briefly, cutoff for YFP ON was determined by thresholding. Cells having YFP intensity greater than the threshold were assigned 'somatic state'. DBSCAN was performed as described earlier and fraction of cells belonging to somatic category was determined. The signal from RFP channel (corresponding to germline state) was not used due to remnant signal from degraded protein that interfered with accurate determination of germ

The MATLAB script was executed from SLURM in supercomputing cluster and all images from the same experiment were processed in the same batch. No manual filtering or images were performed at any stage of this pipeline.

*Measuring mother-daughter cluster size correlation*

The net fluorescence from individual clusters was determined using Fiji and lineages were manually assigned. Custom script written in Python was used to calculate the mother cluster-daughter cluster size correlations. The total size of mother cluster was estimated by summing up the sizes of daughter cluster right after breaking event was recorded.

*Estimating fraction of somatic, germline and mixed clusters*

Flow cytometry files were analyzed using FlowCytometryTools library in Python3.7. The threshold for YFP and RFP channel was determined from the histogram of size (FSC-A) normalized fluorescence values. A cluster was assigned 'Somatic only' type if the YFP intensity was greater than the threshold and RFP intensity was less than the threshold. Similarly, a cluster was assigned 'Mix' type if both the intensity values were greater than the corresponding thresholds. Separate thresholds were determined for different samples and strains.

**Acknowledgements**

We would like to thank staffs, especially Zach Niziolek from Bauer flow cytometry core facility for assistance in FACS and flow cytometry assistance. We would like to thank Christian Hellriegel from Harvard Biological Imaging Center (HCBI)/Zeiss. We extend our sincere gratitude to Dr. and Mrs. Daniel Simmons for awarding Simmons Award to PN for performing advanced microscopy in HCBI.

**Reference**

1 Knoll AH. The multiple origins of complex multicellularity. *Annu Rev Earth Planet Sci* 2011; **39**: 217–239.

2 Brunet T, King N. The Origin of Animal Multicellularity and Cell Differentiation. *Dev Cell* 2017; **43**: 124–140.

3 Staps M, van Gestel J, Tarnita CE. Emergence of diverse life cycles and life histories at the origin of multicellularity. *Nat Ecol Evol* 2019; **3**: 1197–1205.

4 Ratcliff WC, Denison RF, Borrello M, Travisano M. Experimental evolution of multicellularity. *Proc Natl Acad Sci U S A* 2012; **109**: 1595–1600.

5 Koschwanez JH, Foster KR, Murray AW. Improved use of a public good selects for the evolution of undifferentiated multicellularity. *Elife* 2013; **2013**. doi:10.7554/eLife.00367.

6 Libby E, Ratcliff W, Travisano M, Kerr B. Geometry Shapes Evolution of Early Multicellularity. *PLoS Comput Biol* 2014; **10**. doi:10.1371/journal.pcbi.1003803.

7 Koschwanez JH, Foster KR, Murray AW. Sucrose utilization in budding yeast as a model for the origin of undifferentiated multicellularity. *PLoS Biol* 2011; **9**. doi:10.1371/journal.pbio.1001122.

8 Wahl ME, Murray AW. Multicellularity makes somatic differentiation evolutionarily stable. *Proc Natl Acad Sci U S A* 2016; **113**: 8362–8367.

9 Jacobeen S, Pentz JT, Graba EC, Brandys CG, Ratcliff WC, Yunker PJ. Cellular packing, mechanical stress and the evolution of multicellularity. *Nat Phys* 2018; **14**: 286–290.

10 Bozdag GO, Libby E, Pineau R, Reinhard CT, Ratcliff WC. Oxygen suppression of macroscopic multicellularity. *Nat Commun* 2021; **12**. doi:10.1038/s41467-021-23104-0.

11 White MD, Zenker J, Bissiere S, Plachta N. How cells change shape and position in the early mammalian embryo. *Curr Opin Cell Biol* 2017; **44**: 7–13.

12 Kang PJ, Hood-DeGrenier JK, Park HO. Coupling of septins to the axial landmark by bud4 in budding yeast. *J Cell Sci* 2013; **126**: 1218–1226.

13 Ziv N, Siegal ML, Gresham D. Genetic and nongenetic determinants of cell growth variation assessed by high-throughput microscopy. *Mol Biol Evol* 2013; **30**: 2568–2578.

14 Kuranda MJ, Robbins PW. Chitinase is required for cell separation during growth of Saccharomyces cerevisiae. *Journal of Biological Chemistry* 1991; **266**: 19758–19767.

15 Day TC, Höhn SS, Zamani-Dahaj SA, Yanni D, Burnetti A, Pentz J *et al.* Cellular organization in lab-evolved and extant multicellular species obeys a maximum entropy law. *Elife* 2022; **11**: 1–28.

16 Ottoz DSM, Rudolf F, Stelling J. Inducible, tightly regulated and growth condition-independent transcription factor in Saccharomyces cerevisiae. *Nucleic Acids Res* 2014; **42**. doi:10.1093/nar/gku616.

17 King L, Butler G. Ace2p, a regulator of CTS1 (chitinase) expression, affects pseudohyphal production in Saccharomyces cerevisiae. *Curr Genet* 1998; **34**: 183–191.

18 Sbia M, Parnell EJ, Yu Y, Olsen AE, Kretschmann KL, Voth WP *et al.* Regulation of the yeast Ace2 transcription factor during the cell cycle. *Journal of Biological Chemistry* 2008; **283**: 11135–11145.

19 Park HO, Chant J, Herskowitz I. BUD2 encodes a GTPase-activating protein for Budl/Rsrl necessary for proper bud-site selection in yeast. *Nature* 1993; **365**: 269–274.

20 Chant J, Mischke M, Mitchell E, Herskowitz I, Pringle JR. Role of Bud3p in producing the axial budding pattern of yeast. *Journal of Cell Biology* 1995; **129**: 767–778.

21 Lerman RI. a Note on the Calculation the Gin1 Index. *Econ Lett* 1984; **15**: 363–368.

22 Matsuda M, Koga M, Woltjen K, Nishida E, Ebisuya M. Synthetic lateral inhibition governs cell-type bifurcation with robust ratios. *Nat Commun* 2015; **6**: 1–12.

23 Jones ML, Rivett DW, Pascual-Garria A, Bell T. Relationships between community composition, productivity and invasion resistance in semi-natural bacterial microcosms. *Elife* 2021; **10**: 1–25.

24 Dolega ME, Monnier S, Brunel B, Joanny JF, Recho P, Cappello G. Extra-cellular matrix in multicellular aggregates acts as a pressure sensor controlling cell proliferation and motility. *Elife* 2021; **10**: 1–33.

25 Stassen OMJA, Ristori T, Sahlgren CM. Notch in mechanotransduction – from molecular mechanosensitivity to tissue mechanostasis. *J Cell Sci* 2021; **133**. doi:10.1242/jcs.250738.

26 Sancho R, Cremona CA, Behrens A. Stem cell and progenitor fate in the mammalian intestine: Notch and lateral inhibition in homeostasis and disease. *EMBO Rep* 2015; **16**: 571–581.

27 Toda S, Mckeithan WL, Hakkinen TJ, Lopez P. Engineering synthetic morphogen systems that can program multicellular patterning. *Science (1979)* 2020; **331**: 327–331.

28 Ralser M, Kuhl H, Ralser M, Werber M, Lehrach H, Breitenbach M *et al.* The Saccharomyces cerevisiae W303-K6001 cross-platform genome sequence: Insights into ancestry and physiology of a laboratory mutt. *Open Biol* 2012; **2**. doi:10.1098/rsob.120093.

29 Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* 2007; **2**: 31–34.

30 Wach A, Brachat A, Pöhlmann R, Philippsen P. New heterologous modules for classical or PCR‐based gene disruptions in Saccharomyces cerevisiae. *Yeast* 1994; **10**: 1793–1808.

31 Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science (1979)* 1999; **285**: 901–906.

32 Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S *et al.* Functional profiling of the Saccharomyces cerevisiae genome. *Nature* 2002; **418**: 387–391.

**Figure captions**

**Figure 1** **A dynamic network-based approach to examine the mode of breaking in yeast multicellular clusters** A) Timelapse for growth and fracture of multicellular yeast clusters over several generations and a simulation using MultiCellNet. B) Putative models for growth and fracture for the clusters C) Experimental approaches to regulate parameters of interest, i.e., mode of budding (kissing number), link breaking rate and doubling time D) Comparison of size distribution derived from simulation vs experimental estimation

**Figure 2 A high-throughput pipeline to estimate the number of cells in each cluster** Clusters are grown in conditions of interest, squeezed between a coverslip and microscope slide at optimized conditions and imaged using light microscopy. The cells are segmented using an RCNN-based segmentation pipeline. Centroid coordinates are used to perform density-based spatial clustering of applications with noise (DBSCAN), and the number of cells per group for hundreds of clusters is estimated.

**Figure 3 Steric-limited Age-dependent (SA) model recapitulates experimental cluster size statistics** A) 3-dimensional rendering of multicellular cluster with fully induced Bud4p expression and location of budding septum B) Cluster size distribution statistics for putative models and experimental data for low kissing number (. C) Distribution obtained from models (SA and SR) and experimental measurements for full induction D) 3-dimensional rendering of multicellular clusters with no Bud4p induction E) Cluster size distribution statistics for putative models and experimental data for high kissing number ( F) Distribution obtained from models (SA and SR) and experimental measurements for no induction. G) Cluster size distribution for different levels of Cts1p controlled through beta-estradiol induction. H) Relation between mean and standard deviation of cluster size distribution obtained in (G) for two independent biological replicates. The fit is to an Erlang's distribution with shape parameter, k=2.5.

**Figure 4 Live tracking of cluster growth and fragmentation shows correspondence with SA model characteristics** A) The number of cells per cluster proxied by the total fluorescence units collected and division is marked by the separation of clusters B) Distribution of the ratio of the mean number of cells per cluster in the mother to the daughter cluster. C) Comparison of experimental and simulation quantiles for distribution presented in B. D) Lineage tracking of cluster division over several growth-fracture cycles.

**Figure 5 A rheostat-like model operates to control cluster size distribution** A) A hypothesis that a cellular growth rate higher than the link breaking rate yields clusters with a higher number of cells per cluster. B) Clusters with uninduced Bud4p show an increase in mean cluster size with an increase in cell division rate C) Simulations suggests a model in which growth rate sets a threshold for mean breaking time to affect the mean cluster sizes. The mean cluster sizes reach a maximum value independent of cell division rate or link breaking rate D) Experimental data for change in the mean number of cells per cluster with increasing growth rate and chitinase (Cts1p) activity. E) Images for cluster size distribution for increasing growth rate and chitinase activity.

**Figure 6 Kissing number () affects compositional homeostasis in differentiating clusters** A) A hypothesized scenario in which breaking off from a differentiated lineage leads to the formation of a sterile cluster. In this differentiation system, the germline can be converted to a somatic state, not vice-version. The system controls the switching rate from germline to soma and their relative growth rates. B) The SA-based MultiCellNet suggests increasing the kissing number reduces the fracture of sterile clusters while increasing the clusters with both cell types (mixed) presence. The effect is further improved by reducing the relative growth advantage of the somatic cell type compared to the germline cell-type. C) Germline-Soma-Mix composition for clusters with the increased kissing number and relative growth rates determined experimentally. D) Images for clusters with two cell types for various conditions are shown in (C).

**Figure 7 Composition of differentiating cluster is predictable through MultiCellNet and noise in which is minimized at higher kissing number** A) Search space for parameters that can reduce noise or heterogeneity in cluster size composition in differentiating multicellular clusters B) MulticellNet simulations based on SA model predicts the mean fraction of somatic cells in clusters can be controlled by altering the net switching rate (= ) where is the state-switching rate and is the relative growth of somatic with respect to germ cells. C) Experimental measurements of mean somatic fraction in clusters with different genetic backgrounds. The switching rate was controlled by beta-estradiol induction of Cre recombinase and relative growth by cycloheximide. D) Relation between Gini coefficient (measure of cluster composition heterogeneity) and mean somatic fraction for various backgrounds E) Distribution of Gini coefficients for all different values of net switching rate.

**Supplementary Figure 1** A) Evolution of cluster size distribution over 48 h growth for four fractions sorted from clusters with Cts1p shutdown. B) Correlation between mean cluster size measured by counting cells per clusters and multisizer based volume measurement. C) Correlation between mean cluster size measured by counting cells per clusters and flow cytometry

**Supplementary Figure 2** A) Estimated number of clusters required to test for a certain effect size predicted by Cohen's d. B) CDF for cluster size distributions for uninduced Bud4p in *ace2∆* background for two independent replicates C) CDF for cluster size distributions for uninduced Cts1p for two independent replicates D) CDF for cluster size distributions for fully partially induced (8 nM) Cts1p for two independent replicates

**Supplementary Figure 3** Example images of Bud4p uninduced clusters in *ace2∆* background A) DIC image B) DAPI C) Mask generated from RCNN segmentation D) DBSCAN based assignment of cells to a cluster

**Supplementary Figure 4** A) Screen for monotonic and significant increase in cluster size distribution using strains with inducible control of cell-shape properties B) Screen for monotonic and significant increase in cluster size distribution using strains with inducible control of budding pattern

**Supplementary Figure 5** Cluster size distribution generated from NSA (A) and SO (B) based MultiCellNet model for high and low kissing number C) Change in mean cluster size through Cts1 expression on induction with beta-estradiol

**Supplementary Figure 6** Two independent time-traces of growth and fracture in multicellular clusters   
  
**Supplementary Figure 7** A) Images of clusters with uninduced Bud4p expression in *ace2∆* background for two different glucose concentrations B) Growth curves for clusters with uninduced Bud4p expression in *ace2∆* C) Growth curves for clusters with uninduced Cts1p expression and with partial (8 nM) induction of Cts1p D) Flow cytometry based cluster size measurements for growth in two different concentrations of glucose and galactose and for various levels of Cts1p induction.

**Supplementary Figure 8** A) Scatterplots for determining fraction of cluster comprised on purely somatic or germline or having presence of both for various backgrounds.

**Supplementary Figure 9** A) Correlation between two independent biological replicates for three strains used for estimating cluster composition B) Relation between mean somatic fraction and coefficient of variation for various strain backgrounds C) Model predicted relation between mean somatic fraction and Gini coefficients for altering kissing number D) Distribution of Gini coefficients for two different kissing number E) Relation between mean somatic fraction and coefficient of variation for various strain backgrounds for two different link breaking rate F) Distribution of Gini coefficients for two different link breaking rate.