Title
Finding Organization in Randomness: Using Statistics to Explore the Underlying Organizational Character of Intracellular Geometry

Permalink
https://escholarship.org/uc/item/2s245293

Author
Apte, Zachary Schulz

Publication Date
2012

Peer reviewed|Thesis/dissertation
Finding Organization in Randomness:
Using Statistics to Explore the Underlying Organizational Character
of Intracellular Geometry

by

Zachary Schulz Apte

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
ACKNOWLEDGEMENTS

I would like to express my profound appreciation for my PhD advisor Dr. Wallace Marshall who has been a brilliant scientific mentor, an excellent intellectual foil, and amazingly supportive during my time at UCSF. Among the many ways that Wallace has supported me, he has been extremely flexible in my approach research ideas - letting me choose my own path forward. Further, he has consistently encouraged me to peruse opportunities both inside and outside the academic environment. Thank you Wallace.

As well I want acknowledge the rest of my academic committee. Dr. Joseph DeRisi has also been an amazing mentor and an inspiration. Dr. Andrej Sali’s support from helping me with the writing of my NSF Fellowship Proposal through my qualifying exams has been most appreciated.

All my labmates in the Marshall lab have contributed to my education and my research and I want to acknowledge them all. In particular, I wish to acknowledge Dr. Susanne Refelski and Dr. William Ludington. Both of them have both facilitated my graduate research and education – consistently taking the time to give me excellent feedback and advice.

I want to thank my iPQB classmates – the cohort who started the PhD program with me – for excellent companionship and creative collaborations. In particular I want to acknowledge Laurens Kraal, Aram Avila-Herrera and Peter Cimermančič who have read every academic document that I have written during my time at USCF and have spent months teaching me the basics of biochemistry, cell biology and immunology.
Last but not least I want to thank my family - my parents, Mike and Ursula, for their unerring support thought my education.
ABSTRACT

Intracellular organization – organization of organelles and other compartments inside cell-organization is a complex – and the processes that control it are poorly understood. The research contained within this document is an approach to making this problem more tractable – development of statistical methods to compare changes in intracellular organization between populations of cells.

Further, this work is both experimental and theoretical. The organization of the chloroplast micronucleoids was examined within populations (>1000 cells each) of four strains of Chlamydomonas reinhardtii. Two wild type strains (cc124, cc125) and two cell geometry mutant strains (asq2, bld10) were examined with this method. As well, Monte Carlo simulations of organization were conducted utilizing the same parameters as the real cells. Two-dimensional images of objects were segmented to identify individual objects, and the locations of internal points of interest were recorded.

This method can discriminate organizational mutants from wild type statistically – correctly identifying the known organizational mutants bld10 and asq2 as being significantly different from wild type. As well simulations indicate that the method is broadly sensitive to differences in the type and strength of organization within cells. Looking forward to the future, the sorts of quantitative approaches developed in this paper could well be used to determine what genes are involved in intracellular organization.
# Table of Contents

**Chapter 1.**

Introduction .............................................................................................................. 1

**Chapter 2.**

An Information Theoretic Discussion of Spatial Organization ......................... 5

**Chapter 3.**

Comparing The Level of Intracellular Organization Between Cells – a Statistical Method ........................................................................................................ 17

**Chapter 4.**

Summary and Perspectives ..................................................................................... 47
List of Tables

Chapter 3.

Table 1 .................................................................................................................. 46

Table 2 .................................................................................................................. 46
List of Figures

Chapter 3

Figure 1 ........................................................................................................ 39
Figure 2 ........................................................................................................ 40
Figure 3 ........................................................................................................ 41
Figure 4 ........................................................................................................ 42
Figure 5 ........................................................................................................ 43
Figure 6 ........................................................................................................ 44
Figure 7 ........................................................................................................ 45
Chapter 1. Introduction
There are some things that we just know and that are quite hard to measure. Like pornography [1]. If you see it - you know. Physical Organization is a bit like that. Take, for example, laundry. Is your laundry more ordered bunched-up in the basket, or folded? If you said more ordered folded (don't be funny now) how much more ordered? How would you measure it? When we go on to tackle the much more challenging subject of complex physical systems and emergent order, measuring the relative physical organization of a given physical state becomes incredibly important.

In recent years, there has been intense interest in systems biology [2]. Systems biology is the study of how the interactions of the chemical components of life, in large part proteins and genes, generate the complex and highly ordered structures and interactions observed in living organisms [2]. As with the laundry example, even a child can see that cells are extremely organized. However, to rigorously study how organization emerges in these complex physical systems we need to clearly define what we mean by physical organization, and understand how to measure it.

In other fields, organization is approached in a mathematical and systematic way as a generalization of the concept of entropy (disorder) from statistical thermodynamics [3]. For example: in the fields of computer science, cryptography and even genetics, it is possible to construct a metric of how organized a sequence of symbols is - information [4,5]. Recently, in physics, these techniques related to information have been applied to the hard-sphere packing problem [6]. The hard-sphere packing problem is concerned with understanding the configurations a set of spheres can be packed into.

**ENTROPY OF POWDERS**
A series of recent papers have explored how to measure the relative organization of these configurations [6,7,8]. The methods they use to measure the packing of hard spheres are fundamentally based on concepts developed in the seminal paper exploring the organization of spilled powers on surfaces [9]. Essentially, the method is taken in analogy to thermodynamic entropy in statistical mechanics. Where thermodynamic entropy measures how evenly energy is distributed between states, Edwards entropy measures how evenly unoccupied volume is associated with each powder grain. However, Edwards never defines how to measure a local volume associated with a powder grain, instead avoiding the measurement problem altogether [9]. Because the hard-sphere packing problem needs to accurately measure the entropy of different well-defined states they took on this problem - showing that the local volume in Edwards entropy is equivalent to the volume taken up by the volumes defined by the Voronoi tessellation of the centers of the spheres [6].

**VORONOI TESSELLATION**

Voronoi tessellation is the subdivision of space on the basis of the location of a set of points. A Voronoi tessellation is assembled in three-dimensions by intersecting planes normal to, and through the mid-point of, line-segments defined between local points via Delaunay triangulation [6]. These planes intersect with one-another to create closed Voronoi-volumes for surrounded points, or infinite Voronoi-volumes for boundary points. However, if the points are contained within the boundaries of a closed cell, the Voronoi tessellation can simply be bounded at the surface of the cell, subdividing the cell and leaving no infinite Voronoi-volumes.
CITIATIONS

1. Justice Potter Stewart, concurring opinion in Jacobellis v. Ohio 378 U.S. 184 (1964)


Chapter 2. An Information Theoretic Discussion of Spatial Organization
INTRODUCTION

Through much of the history of biological science a major theme has been the attempt to understand the reason that living organisms are much more complex in their organization then the rest of the natural world. While we now have a good understanding of many of the processes involved in the genetic and chemical complexity of life, the mechanisms behind the organization of many of the awe-inspiring macro-scale structures of living organisms are not yet understood. This is due, in part, to the fact that there is no universal measure of how organized a set of extended physical objects in space is. Over this past century, breakthroughs in the quantification of entropy in physics and chemistry and information theory in genetics have lead to large paradigm shifts in our technology and our understanding of those fields. In this paper I discuss how to generally apply the immense power of these statistical techniques to the organization of macro-scale structures.

BACKGROUND

Sadi Carnot was the first to put forward that idea that the motive power of heart is derived from the natural tendency of heat to flow from a hot body to a cold body until they reach equilibrium [1]. This principle is enshrined in the second law of thermodynamics, which states that physical systems maximize their entropy, or disorder, until they reach equilibrium. The development of statistical thermodynamics by Ludwig Boltzmann [2] codified this principle of entropy in terms of absolute statistical entropy,

\[ S = -k_b \sum_i p_i \ln p_i \]  

[Eq. 1]
where $p_i$ is the probability of a given microscopic state occurring in the thermodynamic system in the context of the macroscopic state of the system.

Claude Shannon extended this work to apply to symbolic communication in 1948 [3].

\[ I = -\log p_i \quad [\text{Eq. 2}] \]

Where $I$ is the self-information, also known as surprisal, of receiving a particular symbol out of all possible symbols. Then analogous to [Eq. 1], the total information of a message is given by the probability-weighted sum over the surprisal of each symbol,

\[ H = -\sum_i p_i \log p_i \quad [\text{Eq. 3}] \]

The key concept is that analogous to entropy, the total information content in a communication is determined by assembling the probability of each symbol $p_i$ that appearing in the communication. The unit of information described by Shannon entropy is determined by the base of the logarithm in [Eq. 2, Eq. 3]. Piggybacking on the mathematical framework developed by Boltzmann, this concept of information has been extended and used extensively in many fields including data encryption, computer science and genetics.

The power of information theory, as a tool, is derived from its ability to help us discern order from randomness. For example, the Kullback–Leibler Divergence, can be
used to calculate the gain in information implied between two probability distributions [4],

\[ D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \]  

[Eq. 4]

If a probability distribution, \( Q \), corresponding to the completely disordered state is known then using this tool the amount of information gain over the random state contained in an ensemble distribution can be calculated.

The amount of information encoded in physical systems was, of course, first discussed by Boltzmann [2], however that information was only applicable to the unobservable microstates of a larger physical ensemble. While problems such as the ordering and entropy of packed spheres have been addressed [5], the more general problem of the ordering and entropy of any general physical ensemble has not been discussed in the literature. However, it is intuitively clear that the information encoded in a system is related to the probability of that state occurring.

Methods to calculate the probabilities of spatial distributions of randomly distributed points in two-dimensions were first pioneered by ecologists studying the location distribution of trees in forests [6]. Their method was geared towards using the mapping of a small subsection of forest or other area. They calculated the probability density function of nearest neighbor distances in a small subsection of their sample based on an average density derived from the number of points in their sample. The limited themselves to nearest neighbors because they were attempting to eliminate the edge effects created by their limited sample area. The work in [6] was later extended to measure the probability distribution of nearest neighbors between circles in two-
dimensions rather than points [7]. Posthumously, Francis Evans’ derivation extension of the complete-spatial randomness of nearest-neighbor distances [6] to the $k$-dimensional generalization [8] was published by his coauthor.

In this paper I derive and discuss methods related to those mentioned above, that are, to my knowledge new. These methods are broadly applicable to measuring the self-information contained within macro-scale distributions of extended physical objects within a closed volume. As well, they can be applied to measure the information-gain over complete-spatial randomness found within populations of closed volume spatial distributions such as single-celled organisms.

RESULTS

1. The probability density function for the $n^{th}$ point particle within a space $S$ to be a distance $r$ away from another point contained within a closed space $S$ is,

$$q_n(V_{sp}(r)) = \frac{d}{dr} \left( \frac{\rho V_{sp}(r)^n}{n!} e^{-\rho V_{sp}(r)} \right)$$

Where $V_{sp}(r)$ is the volume of the space created by the intersection of the space $S$ and a sphere $P$, centered on the reference point with a radius $r$. $\rho$ is the density of points in space, generally defined to be the ratio of the number of points in $S$ and the volume of the space $S$.

2. The probability distribution function for spheres of different sizes contained within a closed space $S$ is closely related to [Eq. 5], although it must account for the size of the
spheres: \(R_A\) being the size of the source sphere and \(R_B\) the size of the target sphere. \(C\) is the number of classes of spheres that there are within the distribution, \(S_i\) is the probability of randomly selecting the \(i\)th sphere, and \(R_i\) is radius of the \(i\)th sphere

\[
p_{i}(r,A,R_i,B,R_B,p,P,C) = \left(\frac{\partial}{\partial r} f_i(r,A,R_i)\right) \left(\frac{1}{1 - f_i(R_A + R_B\|A,R_i)}\right) \left(\sum_{j=0}^{\infty} S_j f_j \left(\frac{R_j}{R_A + R_B\|A,R_i}\right)\right)
\]  \[\text{Eq. 6}\]

3. The probability of a whole ensemble, relative to a single point, \(i\), is given by,

\[
P_i = \prod_{j=0}^{N} P_{Q_{ij}}(r_{ij},i,j,R_j,P,C)
\]  \[\text{Eq. 7}\]

Where \(Q_{ij}\) is the distance ordered index of the position of the \(j\)th point relative to the \(i\)th reference point, and \(r_{ij}\). The scope of the ensemble, the points over the index \(j\), can be changed to reflect a single class of points, or a subpopulation of points relative to a single point without biasing the complete spatial randomness.

Principally these equations can be used with [Eq. 2, Eq. 3, and Eq. 4] to calculate the information contained within in an ensemble. However, this is not a complete description of organization because it is only true in reference to the reference point. A more complete description can likely be derived by considering multiple starting points, but the covariance of the probabilities between the different starting points must to be taken into account.
DERIVATION

Given a set of extended physical objects at rest in an enclosed space, in the absence of organizational rules, we would expect a random distribution of those objects. What this means is that for any given object, before we observe it, we will perceive that it is equally likely to be in any location as any other. As is done in the literature [6,7,8], we can first consider the distribution of points in space. If in our enclosed space \( P \), we have \( N \) points, we have a density of points \( \rho = N/V_P \). As we are considering the distribution of points to be uniform within the volume of the space \( P \), we can describe the probability of finding \( n \) points within a subvolume, \( V \), of the space as a Poisson process [8]. This means that we have a Poisson distribution with \( \lambda = \rho/V \) giving,

\[
f_n(\rho V) = \frac{(\rho V)^n e^{-\rho V}}{n!} \quad \text{[Eq. 9]}
\]

If we now want to investigate the relationship between points in our space, we can consider the distance, \( r \), between points in spherical coordinates. If we want to know the probability of the \( n \)th point away from an object occurring within a distance \( r \), we can calculate the volume, \( V_{SP} \), defined by the intersection of a sphere of radius \( r \) centered on our point and the boundaries of the space \( P \) and apply it to [Eq. 9]. Then to calculate the probability density function anywhere within the space we can utilize the fact that,

\[
f_n(\rho V) = \int q_n dr \quad \text{[Eq. 10]}
\]
Where \( q \) is the probability density function as a function of the radius. To calculate \( q(r) \) within the space, we can take the derivative,

\[
q_n(r) = \frac{d}{dr} \left[ \int q_n dr \right] = \frac{d}{dr} \left[ f_n (\rho V_{sp}(r)) \right] = \frac{d}{dr} \left[ \frac{(\rho V_{sp}(r))^n e^{-\rho V_{sp}(r)}}{n!} \right]
\]  

[Eq. 11]

This equation then gives the probability density of finding the \( n \)th point a distance \( r \) away from any reference point. However, this only applies to points. With non-overlapping extended objects the situation is more complicated because the presence of each object excludes volume and thus possible states for other extended objects.

While it is clear that one can construct an arbitrary probability density function for the closest distance between arbitrary shaped objects, here we will consider the simplest case of spheres. Analogous to what is done in [7], we can see that in a point distribution, far away from the boundaries of the space, the probability that no points will both fall within a distance of \( d \) of each other is given by,

\[
q_0(d) = e^{-\frac{4\pi \rho d^3}{3}}
\]  

[Eq. 12]

This means that if we say our originating point is a sphere of radius of \( R_A \) and the second point is a sphere of radius \( R_B \), the probability of \( r < (R_A + R_B) \) goes to zero, and everywhere else is increased by a factor of,

\[
1 - q_0(d) = 1 - e^{-\frac{4\pi \rho (R_A + R_B)^3}{3}}
\]  

[Eq. 13]
Because unlike [7], we are looking at the $n$th point, we also have to reduce the total probability of the second sphere, $B$, of being anywhere by its probability of overlapping with any other point. We can assemble this probability density by taking the sum over every class of sphere in the ensemble of the product of the probability of any point being of that class times the probability density of that point not being within overlapping distance,

$$\sum_{j=0}^{C} S_j \left( e^{-\frac{4\pi \rho (R_B + R_j)^3}{3}} \right) \quad \text{[Eq. 14]}$$

If we also recognize that using spheres means that we cannot place a point within a distance $R_B$ of the boundary of the space, we can introduce a new boundary $V_{PR}$ a distance $R_B$ within the boundary of the space $P$. If we intersect $V_{PR}$ with a sphere as in [Eq. 11] we can create a new volume $V_{SPR}$. Using this volume with [Eq. 11], [Eq. 13], and [Eq. 14] gives,

$$p_n(V_{SPR}(r,R_B)) = \left( \frac{d}{dr} [\rho V_{SPR}(r,R_B)]^{-1} \right) \left( 1 - e^{-\frac{4\pi \rho (R_B + R_j)^3}{3}} \right) \left( \sum_{j=0}^{C} S_j \left( e^{-\frac{4\pi \rho (R_B + R_j)^3}{3}} \right) \right) \quad \text{[Eq. 15]}$$

However this is limited to cases where both $A$ and $B$ are far from the boundary of the space. To write the correction for this simply requires us to write the $V_{SPR}$, the volume of the intersection, of the sphere $\Phi$ as a function of $r$, the radius of the sphere, the location,
A, of the center of the sphere, and the space \( P_R \) defined as being a depth \( R \) away from the inside of the space \( P \),

\[
\gamma(r, A, R) = V[\Phi(r, A) \bigcap P_R] \quad \text{[Eq. 16]}
\]

Using the same notation as above, this gives [Eq. 6],

\[
p_i(r, A, R_i, B, R_j, \rho, P, C) = \left( \frac{d}{dr} \left[ f \left( p_i(r, A, R_i) \right) \right] \right) \left( \frac{1}{1 - f_i \left( p_i(r, A, R_i + R_j) \right)} \right) \sum_{i=0}^{c} s_i \left( p_i(r, A, R_i + R_j) \right) \quad \text{[Eq. 6]}
\]

Using this equation from a single sphere \( i \), to every point within the larger space allows us to construct the probability of that arrangement in space,

\[
P_i = \prod_{j=0}^{N} p_{Q_0} (r_j, i, R_i, j, R_j, \rho, P, C) \quad \text{[Eq. 7]}
\]

Although [Eq. 6] nicely describes the probability of the \( n \)th single sphere randomly being placed at a distance \( r \) from a single other sphere, it doesn’t relate how one class of randomly distributed spheres relates to another class of randomly distributed spheres.

CONCLUSION
While the equations derived in this paper are heavily dependent on the boundary conditions of the space, and are thus likely to be no further generalizable, they are appropriate for use in a computational context. Simple algorithms for the calculation of volumes and their derivatives makes it no great feat to calculate the spatial information of any organization or set of organizations of spheres. While the calculation of the information contained in a space containing arbitrarily shaped objects is beyond the scope of this paper, we recognize that it simply introduces another free parameter having to do with the angular relationship between the objects, and that derivation is likely to differ from this one only in a few places.

The complexity of calculating the entire ensemble probability with these methods necessitates a less cumbersome method to apply to real-world problems. Part of the reason this difficulty arises with these methods is because they use distance as the fundamental statistic. Ultimately the above methods were abandoned by the author for methods using area as the test statistic.

CITATIONS
5. Shahinpoor, “Statistical Mechanical Considerations on the Random Packing of

6. Clark, et al, “Distance to Nearest Neighbor as a Measure of Spatial Relationships in

7. Simberloff, “Nearest Neighbor Assessments of Spatial Configurations of Circles rather
Than Points”, 1979, Ecology, vol. 60, pp. 679-685

8. Clark et al, “Distance to Nearest Neighbor as a Measure of Spatial Relationships in
Chapter 3. Comparing The Level of Intracellular Organization Between Cells – a Statistical Method (in submission to PNAS)
ABSTRACT

Systems level approaches to analyzing complex emergent behavior require quantitative characterization of alterations of behavior on both the micro and macro-scale. Here we consider the problem of cellular organization and describe a statistical methodology for quantitative comparison of the internal organization between different populations of similar physical objects, such as cells. This comparison is achieved with several steps of analysis.

Two-dimensional images of objects are segmented to identify individual objects, and the locations of internal points of interest, such as organelles or proteins, are recorded. In order to define the configuration of internal points in each object, the individual objects are subjected to bounded Voronoi tessellation - subdividing the bounded area of the object into sub-areas determined by the locations of the internal points of interest. The variance of the sub-areas is calculated, and a Monte-Carlo simulation is used to determine the p-value of the variance of each object in a bounded Poisson-Voronoi null model. These p-values are binned on the basis of the number of points of interest and the size of the object. The bins are sampled uniformly from each population, and the sets of p-value are compared with a Kolmogorov–Smirnov (KS) two-sample test to yield a metric for similarity in degree of organization.

We applied this methodology to ask whether centrioles play a role in global cellular organization, using the green alga *Chlamydomonas reinhardtii* as a model system. Mutant strains with known alterations in centriole number, structure, and position (asq2 and bld10) were used. The KS test-statistic of a comparison between the mutant-populations and wild-type cells showed that both strains had a dramatic difference in the
degree of organization compared to wild-type strains. These computational and experimental results confirm prior observational studies and support the idea that centrioles play a role in generating or maintaining global cellular organization. Our results confirm that this method can be used to sensitively compare the extent and type of organization within cells.
INTRODUCTION

A major outstanding problem in basic biology is how cells generate and regulate their three-dimensional geometry on the molecular level [1]. In addition to being an interesting fundamental science question there are clinical implications as well. In development, differentiation of stem cells into distinct functional cell types is accompanied by characteristic changes in cellular organization [2]. As well, the disruption of cellular organization (dysplasia) is a major hallmark of cancer and the basis of cytopathology [3]. The biochemistry canon presumes cell organization is mechanistically generated from molecular networks and molecular self-assembly [4]. While entire genomes have been sequenced and genome-wide molecular-interaction maps exist for model organisms [5], it remains unclear which molecules regulate the intracellular organization or how they do it [1].

Traditionally, cell organization has been investigated qualitatively [6] or using quantitative techniques specifically designed for a particular system and question [7]. However, cell organization is important to the function and survival of cells [1] so it is likely that non-lethal knockouts of genes pertaining to cell geometry will often create subtle phenotypes that are only distinguishable statistically by considering large numbers of cells. Further, hundreds or thousands of genes may directly or indirectly affect the active organization of many subcellular structures. Thus, connecting the vast amount of molecular data with the complex phenomena of cell-organization requires a rapid, statistically robust and scalable approach to discriminating between different types and levels of organization [7,8]. Further, many structures within cells are thought to arise
through self-organization – but detecting self-organization first requires the ability to quantify organization so that we can ask if it is increasing during a given process.

Cell Organization

Cell organization is defined as the characteristic positioning of organelles within the cell body [6]. Historically, cell organization has been approached qualitatively from two different directions - cell polarization [6] and cell patterning [7]. Polarization generally refers to asymmetries generated within a cell; for example, the determination of a division plane and the localization of the spindle-pole bodies to opposite ends of a cell during division [6]. On the other hand, cell patterning concerns the localization of organelles to specific sub-cellular locations. For example, centrioles are generally localized into pairs in most eukaryotes. The unicellular green alga Chlamydomonas reinhardtii has been used to identify some of the molecules responsible for linking centrioles together in pairs [8, 9]. While both these concepts about organization have revealed processes that are responsible for intracellular organization, they are products of a mostly qualitative framework for understanding organization and do not allow us to compare the relative degree of organization between cell types. A framework that allows us to compare the relative degree of organization between cell types is critical for asking questions such as whether or not a particular mutation, or alterations in a particular organelle (such as the centriole) leads to more or less order in the cell. While a small number of quantitative studies have been conducted within both these conceptual spaces [8,10], the use of cell-specific polarity markers to determine orientation and organization have made these studies non-general by their very nature.
Measuring Organization

In several fields, organization is approached in a mathematical and systematic way as a generalization of the concept of entropy (disorder) from statistical thermodynamics [11]. For example: in the fields of computer science, cryptography and even genetics, it is possible to construct a metric of how organized a sequence of symbols is [12, 13]. Condensed matter physics has long addressed physical order in bulk by defining order parameters such as the magnetization vector to measure degree of organization.

More recently in physics, the same techniques related to information theory have been applied to the hard-sphere packing problem [14]. The hard-sphere packing problem is concerned with understanding the configurations a set of spheres can be packed into. A series of recent papers have explored how to measure the relative organization of configurations of distributions of spheres [14, 15, 16]. The methods used to measure the packing of hard spheres are fundamentally based on the conceptual framework of Edwards entropy [17]. Subsequent work on the hard-sphere packing problem showed that the local volume in Edwards entropy can be taken as equivalent to the volumes defined by the Voronoi tessellation of the spheres [14].

Voronoi tessellation is the subdivision of space on the basis of the location of a set of points [16]. A Voronoi tessellation is assembled in three-dimensions by intersecting planes normal to, and through the mid-point of, line-segments defined between local points via Delaunay triangulation [14]. These planes intersect with one-another to create closed Voronoi-volumes for surrounded points, or infinite Voronoi-
volumes for boundary points. If the points are contained within the boundaries of a closed cell, the Voronoi tessellation can simply be bounded at the surface of the cell, leaving no infinite volumes. This has the effect of subdividing the volume of the cell on the basis of how evenly spaced out the points are within the volume. Applying these statistical and geometric approaches to cell organization is a logical step towards a general ‘order’ parameter – a parameter quantitatively measuring the level of organization in cells.

**THEORY**

*Theoretical Framework for Measuring Cell Organization*

To begin to examine organization in a broad manner, an appropriate definition of organization must be used. In this case, we are interested in organization at the level of organelle positioning (i.e., how a set of organelles are placed in the body of a cell). In a randomly organized cell any organelle is as likely to occupy any one spot in the cell as any other spot, and we therefore seek a definition of organization that quantifies deviation from this minimally organized state. An organized state can therefore be defined as a spatial-bias to the placement of organelles within the body of the cell. Thus, when we talk about organization, fundamentally what we want to understand is how non-random a cell’s organization is. Statistically speaking, this is equivalent to determining a statistical distance between the distribution of a test-statistic in our null-model (a cell with a uniform random spatial distribution of organelles) and the distribution of that test-statistic in a clonal population of actual cells. Extending this concept, differences in the degree of organization between wild-type and mutant cells could be measured with the statistical distance between the test-statistic in each population.
To conduct a statistical analysis of intracellular organization with this conceptual framework, a relevant parameter, or test-statistic, must be used to determine the organizational state of a cell. In this case, a logical parameter is how non-uniformly the organelles are distributed within the volume of a cell in the statistical limit [17]. We propose that one useful mathematical implementation of such a parameter is the variance of the areas found by Voronoi tessellation of the locations of the organelles [Fig. 1]. Using two- or three-dimensional data describing the locations of organelles and the cell boundary, cells can be mathematically divided up into subareas (or volumes) with Voronoi tessellation. The variance of these areas (or volumes) within a single cell is,

\[ \sigma^2 = \langle a^2 \rangle - \langle a \rangle^2 \]  

[Eq. 1].

Unfortunately, this value changes dramatically with the size of the cell, making it an unlikely test-statistic. However, the null-model is a uniform random distribution in space, so it is scale invariant. Thus the p-value (the probability of getting a value at least as extreme as the value obtained) of a given variance (and thus organelle configuration) in the null model is also scale invariant and directly related to how extreme the organizational state is. This makes the p-value of the variance of areas (or volumes in 3D) for a given organelle-configuration a reasonable choice for a test statistic.

**Measuring Organization and Statistical Distance Between Populations**

Empirical results indicate that the variances of Voronoi areas in the unbounded case are well described with a two-function gamma distribution [18, 19, 20]
Logically extending this to the bounded case, the parameters of the gamma distribution for any given cell boundary and number of organelles can be found to arbitrary precision through Monte-Carlo simulation and fitting the resulting distribution to a gamma distribution. Given the subcellular locations of organelles and the location of the cell boundaries for a population of real cells, we can calculate the set of Voronoi areas for each real cell in a population [Fig. 1]. Then for this particular set of areas we calculate the sample variance as a simple measure of non-uniformity [Fig.1a]. We then perform a Monte Carlo simulation of the distribution of variances within that particular cell under the assumption of spatially uniformly distributed points [Fig.1b]. We then can use the results of the Monte Carlo simulation to determine a p-value for any particular observed variance in a given actual cell [Fig. 1c]. Once the p-value of the variance of each individual cell-configuration is calculated for each cell in a population, the Kolmogorov-Smirnov two-sample test can be used to calculate a measure of the statistical distance between two populations of cells. Sampling both populations such that the number of cells with a particular size or organelle number is equal in both samples allows for direct comparison between the populations [Fig. 2]. The Kolmogorov-Smirnov statistic is interpreted as a measurement of the difference in organizational distribution between these different populations. Because only a subset of values is sampled for each bin, bootstrapping by resampling of the dataset 1000 times allows a bootstrap mean and variance to be estimated. Further, examining the distance ‘spectra’, of the value of the
statistical distance versus cell size and organelle number could provide insight into the type of organization observed.

RESULTS

Monte Carlo Simulation of Intracellular Organization

As is discussed above, we are defining cell organization as a non-random spatial bias to the placement of organelles within the body of a cell. In order to validate this strategy, we first simulated several possible modes of cellular organization [Fig. 3] in order to ask whether our method could distinguish them. First we simulated the completely random case, where organelles are placed without any spatial bias [Fig. 3A]. Following that we created a spatial bias by making 1000 random simulations, and choosing the simulation that had the minimal average interorganelle distance to simulate organelle clustering [Fig. 3B]. Then we created a different spatial bias by making 1000 random simulations, and choosing the simulation that had the maximal average interorganelle distance to simulate active organelle dispersal [Fig. 3C]. Finally we created a different spatial bias by placing points in clusters of two points to simulate paired organelles [Fig. 3D]. Simulations indicated that these different types of non-random organization achieved differently shaped curves – leading to non-zero KS distances [Fig. 3E].

Demonstration of method using mutant cells

In order to test whether the above measure of difference in degree of organization can discriminate cell types with biologically significant organization, we performed statistical analysis on the organization of chloroplast and mitochondrial -nucleoids in
*Chlamydomonas reinhardtii* [Fig. 4]. *Chlamydomonas* is a unicellular green alga whose cells have a highly stereotyped polarized morphology with characteristic positioning of organelles. In a typical Chlamydomonas cell, the pair of centrioles is docked at the cell surface at one end (referred to as the anterior end of the cell) and the Chloroplast consists of a single large cup-shaped structure filling up the posterior half of the cell [21]. The chloroplast nucleoids are being used as a proxy for the distribution of the chloroplast itself throughout the cell interior. We applied our analysis to four Chlamydomonas strains: two wild-type strains (cc124 and cc125) which are of opposite mating types and have slight differences in cell size but are otherwise thought to be virtually identical; *asq*2 mutants [8] which are defective in the protein Tbccd1 [22], in which the mother-daughter centriole linkage is abrogated, daughter centrioles move to random positions on the cell cortex, and the total number of centrioles is variable from cell to cell; and *bld*10 mutants [23] which are defective in the protein Cep135/Bld10p that localizes in the cartwheel of the centriole [24], in which centrioles are reduced to small precursors lacking defined microtubule blades and move to a more central position in the cell, near to the cell nucleus away from the cell surface [8]. Visual observations of the two mutants strains have revealed that other cellular structures besides centrioles are misplaced in these mutants [8] making them likely candidates for mutations that might decrease the level of organization, although this was not possible to determine without a quantitative measure of organization. We therefore asked whether our measure of organization would show a significant increase in disorder in these mutant strains when compared to the two wild-type strains, whose degrees of order we expected to be roughly similar.
The Kolmogorov-Smirnov (KS) distance tests, based on the distributions of p-values for deviation in variance of Voronoi tessellation areas, did in fact find significant differences between wild type populations and known mutants [Fig. 5, Table. 1]. Monte Carlo simulations conducted with the same parameters (shape, size and organelle number) as the real cells also showed that the distances between wild-type cells, cc124 and cc125, were consistent with both populations being drawn from the same Monte Carlo distribution [Fig. 5, Fig. 7]. Further, the distances between wild type cc124 and the mutant populations were much greater than for populations with organelle locations drawn from the same type of spatial distribution [Fig. 7, Table. 2]. This is consistent with cc124 and cc125 being internally organized in a similar way, and bld10 and asq2 being organized in a different way than wild-type each other and the other strains.

One of the major challenges with developing an order parameter approach for cells is that every cell is different, and mutants can have for example different cell shapes than wild-type cells. Because our method employs actual measurements of cell shapes as the starting point for Monte Carlo simulations, we expect it to be less sensitive to such effects than would be a method based on assumptions regarding identical cell shapes. In order to rule out the possibility that systematic differences in cell shape or organelle number between mutant and wild-type cells might be responsible for the apparent difference in organization, we conducted Monte Carlo simulations using wild-type and mutant cells to provide the bounding volume shape, size, and number of organelles, but with organelles randomly placed according to the null model, and compared the results for these null model simulations for mutants versus wild type cells to each other [Fig. 6]. We also used the same observed cell shapes and organelle numbers to repeat simulations using
three different models for organizational bias (as had been previously done for purely theoretical shapes in [Fig. 3]). We found that regardless of the organizational bias model, there was no significant difference between results obtained in mutant cells and results obtained in wild-type cells.

An opposite concern was that difference in cell shape or size might cause enough random variation in a population those differences in organizational degree become hard to detect. In order to rule out such an effect, we compared the results of Monte Carlo simulations with biased organization models using cells from the wild type and mutant populations, with a reference simulation of uniformly distributed points in the wild-type cc124 cell populations [Fig. 7]. The results show that the biased organization is clearly detectable, even when comparing cell shapes drawn from two different mutant or wild-type backgrounds.

We conclude that any systematic differences in cell size or shape, or organelle number, can neither produce a statistically significant difference in our calculated degree of order, nor prevent our detection of at least large differences in order, indicating that our use of actual cell shapes for Monte Carlo simulation combined with binning normalization [Fig. 2] is effectively removing the effects cell size and shape in the comparison.

**DISCUSSION**

*Comparison with other quantitative analyses of cell structure*

Past methods that have quantitatively scored cell structure on a set of multidimensional image parameters have managed to score cells as normal or abnormal [25]. Methods have also been described for quantifying cell shape and comparing the
shapes of different cells [26]. However, these past methods were geared towards recognizing distinct cell morphologies, and were not necessarily intended for the specific purpose of measuring order versus disorder. In this paper we have shown a relative way to quantify the level and type of order in cells, a necessary first step towards treating cells as a branch of condensed matter physics.

**Experimental validation**

Overall, the statistical distance tests provide good correspondence with what has been noted from observational studies. Visual examination of bd10 conveys the qualitative impression that these cells are much less organized than all three of the other strains [24], and similarly that asq2 is intermediately organized in comparison [8]. The fact that our prior visual impression has now been recapitulated in the measurement of the cell organization over very large populations of cells supports the idea that the method is discriminating a real phenomenon [Fig 5].

**Potential Applications**

There are a number of potential applications of this method. For instance, this method could be used to measure degree of dysplasia in cancer cells, for clinical and diagnostic purposes as well as a research tool. As well, this tool could rapidly identify gene functions or drug targets that play a role in cellular organization in a high-throughput fashion. As well, many fundamental questions about cell organization over the cell cycle remain, and this tool could give insight into time dependent organizational changes in populations. Further, differences in protein localization could be examined
quantitatively in super-resolution microscopy techniques such as PALM or STORM. A completely different class of applications arises in attempting to test and refine theoretical or computational models for cellular organization and polarity. In general, the big problem with attempting to compare real cells with simulated cells is that one has to decide which aspects of the real-world cells to consider. This method solves that to some degree by providing a means to benchmark theoretical models using experimental data.

**MATERIALS and METHODS**

*Cell Culture*

Four strains of *C. reinhardtii* cells, cc124, cc125, asq10, and bld10 were obtained from the Chlamydomous Genetics Center [27]. Each strain was grown and maintained in Tris-acetate-phosphate (TAP) media [28].

*Microscopy*

Cells were fixed with 5% glutaraldehyde and stained with DAPI and FITC conjugated Concanavalin-A (FITC-ConA). After fixing the cells were suspended in TAP media, and 10 ul of cell-media solution was mounted on a microscope slide between a standard 22 mm square coverslip and sealed with Vaseline. Slides were imaged using a 20x air lens on a Deltavision deconvolution microscope (Applied Precision, Issaquah Washington). Automated 2D imaging in both the FITC and DAPI channels across large areas of the slide was achieved with Softworx imaging software and an automated microscope stage.
Image Segmentation

Image segmentation was used with the images generated from the FITC channel of the FITC-ConA stained cells to find the cell outlines with the MATLAB Image Processing Toolbox. Each microscope image was broken into sub-regions to isolate each cell [Fig. 4]. Image segmentation of DNA spots in the DAPI channel was achieved using the MATLAB Image Processing Toolbox. First, a threshold was set for each image using the Ridler-Calvard algorithm. Then, intensity peaks in the image were identified using the peaks of connected pixel regions 1.5 times above the threshold [Fig. 4].

Bounded Voronoi Tessellation of a Cell

In two-dimensions, the boundary of the cell is represented by an ellipse and each organelle in the cell is approximated by a point. Voronoi tessellation of each point within the cell with the Multi-Parametric Toolbox in MATLAB [29] defines the region of the plane closest to each point in the set of points. The intersection of this region with the area inside of the ellipse defines the area within the cell that is closest to the each organelle. The variance of the areas defined by the intersection of the Voronoi-facets and the ellipse is then calculated [Fig. 1a].

Null Model of Organization in a Particular Cell

A null-model of the organization of each cell was constructed using Monte Carlo simulation. The same number of points as were segmented in the real cell was randomly placed within the cell boundary. Iterating the Monte Carlo simulation 1,000 times and calculating the variance of the bounded Voronoi tessellation of each generated the null-
model distribution of variances [Fig. 1b]. This distribution was then approximated with a 2-variable gamma distribution using the MATLAB Curve Fitting Toolbox.

**P-value of a Cell’s Organization in the Null Model**

Calculating the p-value of a particular cell is achieved in two parts. First the variance of the Voronoi areas of the real cell is calculated (as above). Then the distribution of the variances of the Voronoi areas is calculated (as above). The cumulative distribution function of the 2-variable gamma function can be directly used to calculate the p-value [Fig. 1c]. In this research the p-values found were below the floating-point precision of normal computer architecture – so the natural logarithm of the p-values was directly calculated instead – allowing the computer to handle subsequent calculations.

**Binning, KS-Distance and Bootstrapping Methodology**

To directly compare two different populations of cells, it is extremely important to control for bias that is introduced by comparing sets of cells with different area distributions or different numbers of organelles. To control this, each population is binned in two dimensions – the total area of each cell and the number of organelles in each cell [Fig. 2]. Corresponding numbers of cells are randomly drawn from each bin from each population to create sets for each population. The Kolmogorov-Smirnov (KS) two-function test is then used to calculate a statistical distance between the p-value distributions of each population set. Because a random subset of the values in each bin is being used to generate comparison sets, this random selection can be iterated to generate
numerous bootstrapped KS distances. This set of bootstrapped KS distances can then be used to calculate the mean distance between the populations as well as the standard deviation of that distance. The mean KS distance obtained in this way is used as our ultimate measure of difference in degree of organization.

*Control Simulations*

Simulations were conducted to test if the comparison between two populations of cells is biased by the size of the cell or the number of organelles in the cell. For each cell in both real populations the parameters (size, shape and number of organelles) were taken. Each set of cell parameters was used to randomly place organelles within the cells. These sets of simulated cells with real parameters are compared using the methods above. The statistical distance of the results of the comparison of these simulated populations can be then used to set a baseline for the size of the statistical distance between any two populations.

*Non-Random Cell Generation*

To directly compare simulations more closely with the real the real experiments, Monte Carlo simulation was used to generate random and non-random datasets using the same parameters, (shape, size and number of organelles) as the populations of real cells. The non-random datasets were created by randomly placing points into an ellipsoid rotated from the ellipse representing each cell. The points x and y coordinates were kept. The non-random simulations were generated similarly – each cell in the “Min” simulation represents the selection of the cell with the minimum average interorganelle distance
from a set 1000 randomly simulated cells [Fig. 3B]. Each cell in the “Max” simulation represents the selection of the cell with the maximum average interorganelle distance of 1000 randomly simulated cells [Fig. 3C]. The organelles within the Point Clumps simulation are placed in close point pairs at random locations within the cell [Fig. 3D]. These simulations were then analyzed in exactly the same manner as the real cells – comparing each simulation a simulation of the same type with the parameters of the cc124 cells. However, the comparison of different types of simulations [Fig. 6] was calculated by measuring the distance from each non-random simulation type to the synthetic cc124 dataset.

ACKNOWLEDGEMENTS

We thank Susanne Rafelski, William Ludington, Aram Avila-Herrera, Peter Cimermančič and Michael Apte for their helpful editing of this manuscript. This work was supported by grant GM077004 from the National Institutes of Health.

CITATIONS


3. Pfau and Chak. Detection of preinvasive cancer cells: early-warning changes in precancerous epithelial cells can now be spotted in situ and Endoscopic detection of
dysplasia in patients with Barrett’s esophagus using light-scattered spectroscopy.

Gastrointestinal endoscopy (2001) vol. 54 (3) pp. 414


FIGURES

Fig. 1. This figure shows a cartoon of the steps taken in the statistical method. (a) The Voronoi tessellation of a single cell and the calculation of the variance of the Voronoi-facets. (b) Monte Carlo simulation is used calculate the distribution of variances for the null-model in a given cell. (c) The p-values for each cell in the null model are calculated using the variance of the real cells and the distribution of variances found by Monte Carlo simulation.
Fig. 2. The p-value of each cell in the experiment is binned using a two-dimentional binning scheme, seperating the cells by size and by number of organelles. This is to ensure that when comparing populations, equal numbers of cells with similer size and number of organelles can be used.
Fig 3. This figure shows examples of different kinds of organization and the distribution of p-values that each type of organization creates. (A) Shows a randomly organized cell, whereas (B) shows a cell with the minimum interorganelle distance from among 1000 random cells, (C) shows a cell with the maximum interorganelle distance from among 1000 random cells and (D) shows points placed in clumps. (E) Shows the frequency versus ln ( p-value ) curves for each type of simulation.
**Fig 4.** (A) An automatically segmented *C. reinhardtii* cell. The cell boundary is indicated in red, and the locations of the organelles, in this case DNA-micronucleoids are indicated with white circles. (B) A transmitted-light micrograph of the same cell. (C) Fluorescence in the DAPI channel, indicating the presence of DNA, in the same cell.
Fig 5. The Kolmogorov Smirnov statistical distance between populations of cc125, asq2 and bld10 cells and a population of cc124 cells (in blue). Identically sized samples were taken from several size and area bins for each cell line. In red, this figure shows the same statistical distance measure but applied to simulated versions of the same four strains. These simulated cells had the same parameters as the wildtype cells, but the locations of their organelles were drawn from a uniform random distribution within an ellipsoid. This result clearly indicates that the real distance between cc125 and cc124 is indistinguishable from statistically identical populations in the same cells. As well, it appears that both asq2 and bld10 are significantly distant from cc124 and cc125, indicating a different underlying organizational distribution.
**Fig 6.** Cell shape and organelle differences do not cause significant differences in apparent degree of organization between mutant and wild-type. The ellipse configuration and number of organelles from real cells were used to generate synthetic datasets (corresponding directly to each real population) using the non-biased and biased simulation methods of [Fig 3]. The distance between each of the different synthetic datasets (for cc125, asq2, and bld10) was measured against the synthetic dataset generated using cell shape and organelle number from wild-type strain cc124 simulated with the same spatial model (random and three different bias models). Results show that when comparing these populations with similar underlying distributions using different cell populations for cell shape and organelle number, the resulting Kolmogorov-Smirnov distance is remarkably similar.
Fig 7. Spatial bias can be detected despite variation in cell size, shape, and organelle number between mutant and wild-type strains. The ellipse configuration and number of organelles from each of the real cells were used to generate synthetic datasets (corresponding directly to each real population) using the non-biased and biased simulation methods of Fig 3. The distance between the synthetic cc124 dataset generated with a random distribution of points and each of the synthetic datasets was calculated. Simulations of uniform point distribution in both population pairs (dark-red) contrast strongly with simulations using uniform distribution in one population and biased distribution in the other – revealing sensitivity to changes in organizational type despite variation in cell size, shape, or organelle number.
<table>
<thead>
<tr>
<th></th>
<th>cc124</th>
<th>cc125</th>
<th>asq2</th>
<th>bld10</th>
<th>number of cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc124</td>
<td>1423</td>
<td>0.0544 (0.0113)</td>
<td>0.0846 (0.0149)</td>
<td>0.1266 (0.0132)</td>
<td>0</td>
</tr>
<tr>
<td>cc125</td>
<td>1842</td>
<td>0.0544 (0.0113)</td>
<td>0.0588 (0.0130)</td>
<td>0.1067 (0.0152)</td>
<td>0</td>
</tr>
<tr>
<td>asq2</td>
<td>2365</td>
<td>0.0846 (0.0149)</td>
<td>0.0588 (0.0130)</td>
<td>0.1123 (0.0177)</td>
<td>0</td>
</tr>
<tr>
<td>bld10</td>
<td>1268</td>
<td>0.1266 (0.0132)</td>
<td>0.1067 (0.0152)</td>
<td>0.1123 (0.0177)</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 1.** The KS two-function test statistic between pairs of cells with the bootstrapped standard deviation in parentheses. The number of cells from each type used in the experiment is shown on the right.

<table>
<thead>
<tr>
<th></th>
<th>cc124</th>
<th>cc125</th>
<th>asq2</th>
<th>bld10</th>
<th>number of cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc124</td>
<td>1423</td>
<td>0</td>
<td>0.0544 (0.0158)</td>
<td>0.0472 (0.0139)</td>
<td>0.0410 (0.0099)</td>
</tr>
<tr>
<td>cc125</td>
<td>1842</td>
<td>0</td>
<td>0.0537 (0.0158)</td>
<td>0.0496 (0.0155)</td>
<td>0.0537 (0.0144)</td>
</tr>
<tr>
<td>asq2</td>
<td>2365</td>
<td>0</td>
<td>0.0472 (0.0139)</td>
<td>0.0496 (0.0155)</td>
<td>0.0471 (0.0129)</td>
</tr>
<tr>
<td>bld10</td>
<td>1268</td>
<td>0</td>
<td>0.0410 (0.0099)</td>
<td>0.0537 (0.0144)</td>
<td>0.0471 (0.0129)</td>
</tr>
</tbody>
</table>

**Table 2.** The KS two-function test statistic between pairs of cells with cell organization simulated with a uniform random distribution within an ellipsoid. The bootstrapped standard deviation is in parentheses. The parameters of the cells used in these simulations were identical to the set of cells simulated in Table 1. The number of cells from each type used in the experiment is shown on the right.
Chapter 4. Summary and Perspectives
The potential for using quantitative methods, such as discussed in Chapters 2 and 3, to understand the physical organization of cells is astounding. Imagine how powerful it will be to apply the statistical and genetic tools in used in bioinformatics to problems of cell organization in high-throughput. There is a wide range of potential applications. In this chapter I outline a few of the applications, and suggest some specific research projects that will be valuable to produce next.

**Some Potential Applications:**

**Developmental Biology**

Developing from a single cell into a human being involves a series of incredible transformations in cell architecture. As cells differentiate, cell organization changes dramatically. In particular – think of the organization of cell types in the human body – neurons, muscle tissue, blood cells, b-cells. Creating this astounding diversity of form with identical genomes is achieved through regulation of gene-expression and chemical signals. To date many of these mechanisms aren’t understood. The statistical method described in this paper will allow measurement of small changes in organization within a population over time. This will enable experiments focused on understanding the time progression of organizational changes during differentiation and identification of what genes are involved.

**Rapid Analyses of Cancer Tissue Biopsies**
In many cases, the way that a doctor determines that a patient has cancer is the examination of a tissue sample by specialized doctors. This is as much an art as a science, with the doctor making the determination from training and experience, not a quantitative method. Using this method with tissue biopsies from cancer could allow for comparison of healthy and unhealthy samples rapidly and inexpensively – assisting doctors in making rapid diagnosis for cancer, and increasing the accuracy as well as usefulness of cancer biopsies as a diagnostic tool.

**Research Tool for High-throughput Organizational Screening**

The tools I have developed examine spatial correlations between many populations of cells with discrete organelles. This will allow for high-throughput automated screening for cell-organization. A number of potential systems that could be examined with this tool are cancer-cell lines, model organisms such as *S. cerevisiae* as well as plant tissues.

**Next Research Steps:**

In this author’s opinion, the most powerful application of this method is the third one – as a research tool to find the genetic determinants of cell organization. With this, *S. cerevisiae* can be used to identify and study organization genes on a large scale. High-throughput screening of the *S. cerevisiae* knockout library will allow comparison of the effects of removing each gene from yeast on overall organization and clustering. Double knockouts of genes found within an organizational cluster can be used to help determine potential organization pathways. This will require several steps:

**High-Throughput Screen in the Yeast-Knockout Library**
To conduct this study, the following step will need to be taken:

1. Tagging of the following organelles with different fluorescent proteins:
   a. Golgi Bodies
   b. Peroxisomes
   c. Lipid Droplets
   d. Cell Wall

2. High-throughput Yeast Recombination into the Yeast-Knockout Library

3. High-Throughput Fluorescence Microcopy of 1000 cells in each population.

4. Statistical Analysis and Clustering to Propose Pathways

5. Double-Knockout strains with same protocol to Determine Pathway

Completion of this future project will set the stage for more detailed molecular mechanistic studies in the future. We will have determined genes likely involved in molecular pathways – investigating the biochemistry of the pathways will be the logical next step. As well as following up on the molecular pathways, there are other applications for the statistical methods I described above for quantitative analysis of cell organization. Further experiments involving cell-organization and transcription, the cell cycle or drugs are obvious. As well, these techniques could have applicability in oncology as a novel method for automated scoring of dysplasia in tissue biopsies.
Publishing Agreement
It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:
I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

_____________________________________   ________________
Author Signature        Date

4/23/2012