GENE EXPRESSION PATTERNS OF ENCAPSULATED MICROBIAL CELLS

BY

ZEESHAN FAZAL

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Animal Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Sandra Luisa Rodriguez-Zas, Chair
Associate Professor Maria B. Villamil
Professor Emeritus Eric Jakobsson
Dr. Chester Brown

Abstract

To design hybrid cellular/synthetic devices such as sensors and vaccines, understanding of how the metabolic state of living cells changes upon physical confinement within three-dimensional matrices is vital. We analyze the gene expression patterns of stationary phase *Saccharomyces cerevisiae* (*S. cerevisiae*) cells encapsulated within three distinct nanostructured silica matrices and relate those patterns to known naturally occurring metabolic states. It was found that the cells for all three-encapsulated methods enter quiescent states characteristic of response to stress, albeit to different degrees and with differences in detail. By the measure of enrichment of stress-related Gene Ontology categories, we find that the AqS+g encapsulation more amenable to the cells than CDA and SD encapsulation. We hypothesize that this differential response in the AqS+g encapsulation is related to four properties of the encapsulating gel: 1) oxygen permeability, 2) relative softness of the material, 3) development of a protective sheath around individual cells, and 4) the presence of glycerol in the gel, which has been previously noted to serve as a protectant for encapsulated cells and can serve as the sole carbon source for *S. cerevisiae* under aerobic conditions. This work represents a combination of experiment and analysis aimed at the design and development of 3D encapsulation procedures to induce, and perhaps control, well-defined physiological behaviors.

We also report on the temporal pattern of yeast gene expression patterns during encapsulation in silica matrices via a cell-directed assembly process, and upon release. Three broad classes of patterns are seen. A major shift in expression patterns is seen upon encapsulation, relative to the beginning stationary state, similar to previously reported stress response. Significant continuing shifts are seen by sampling at different intervals during a one week encapsulation. Upon release from encapsulation and reincubation in growth medium, the cells are in a state significantly different from the state prior to encapsulation and similar to the state during encapsulation. Implications are drawn for the use of encapsulated micro-organism as sensors and effectors, and for the persister state of such organisms.

Ordinarily Gene Ontology (GO) enrichment analysis is subject to an arbitrary threshold for defining significance of enriched classes. In this paper, we consider replacing an arbitrary threshold with F-measure optimization to define the *p*-value that divides “significant enrichment” from

"non-significant". It is found that evaluation of false negatives (essential for computing recall and thus F-measure) requires a heuristic (but reasonable) assumption. We apply F-measure optimization to two sets of genes from different organisms and use Benjamini-Hochberg and random resampling to evaluate the number of false positives. It is found that the uncorrected *p*-value that produces optimum F-measure varies widely from one data set to another. It is also found that all three methods of FDR calculation diverge from each other within a range of uncorrected *p*-values that provide F-measure optimum *p*-values. This study includes in Appendix AI a pipeline for using resampling and F-measure optimization to create lists of enriched GO classes that provide for variable weights of precision and recall.

Table of Contents

Chapter 1: Literature Review

Yeast Stress Responses

The yeast *(Saccharomyces cerevisiae)* is a model eukaryotic unicellular organism and the first eukaryotic organism whose genome was completely sequenced.[1] As a unicellular organism, yeast lives freely in natural environment that put yeast under variety of external stresses. Any environmental condition that challenges the cell to survive or perturbs the normal function of cell is termed as stress.[2] The yeast cells must adjust internal system to handle these external changes to perform regular functions. One of the yeast responses to these changes is provisional pause of normal cellular processes during the reorganization of internal system.[2] There are some stress conditions that have been studied experimentally like growth at suboptimal temperature, growth at limited nutrients supply, DNA damaging or toxic agents, growth in a medium of different osmolarity and others. Yeast's genomic response is unique to the features of environment. Scientists carried out yeast comparative analysis of stress responses to varying kind of conditions. These experiments uncovered that there are approximately 900 genes whose expression was changed upon different environmental stresses.[2] These 900 genes were named as Environmental Stress Response (ESR) genes. Environmental Stress Response (ESR) genes were classified into two categories depending upon the expression profiles. Category one contains 300 upregulated genes in response to stresses and these genes were referred as induced genes while category two contains 600 downregulated genes and these genes were referred as repressed genes. It was noted that changes in ESR genes expression were directly proportional to the magnitude of environmental stress. Authors verified this correlation by carried out dose dependent experiment in which they exposed yeast cells to extreme environmental change and observed greater change in ESR gene expression while exposure to delicate or less extreme environmental change led to small change in ESR gene's expression.[3] Most of genes whose expression was downregulated were involved in ribosome synthesis and processing, transcription, translation, and RNA processing. These downregulated processes are energy consuming processes and lower synthesis of these processes will help cell to conserve energy for important cellular functions. However, genes whose expression went upregulated were genes involved in carbohydrate metabolism, autophagy, DNA damage response, and oxidative stress responses.[2] This is very much important for cells to cope

with these environmental challenges for its survival; However, coming back to normal state after adopting to the new environmental is also very critical for cell's survival. To understand this later aspect, further studies needed to be carry out. A comprehensive understanding of yeast's genomic responses under stress will help to grasp the connection between stress responses and aging, identification of new antibiotics, and deeper understanding of evolutionary importance of yeast's stress responses.[3]

Persister Cells and Drug Tolerance

Persister cells were first discovered by Bigger by analyzing the mode of penicillin action.[4] Bigger observed that penicillin broke down the growing *Staphylococcus* spp; however, some small sub-population of bacteria who was not antibiotic resistant was survived. Bigger proposed this small population of surviving cells was non-growing or slow growing, non-dividing and dormant.[4] It was recently discovered that persister cells were not only present in bacterial population but also found in some small population of yeast such as *Candida Albicans* cells who were survived in high concentration of antifungal agents. Microbial population forms persister cells as one mode of survival and these persister cells have same genotype but phenotypic variants of large population, which are multidrug tolerant.[5] Tolerance is the way of cells to evade killing by antibiotic or antifungal agents by altering the target or not synthesizing the target required by antibiotic or antifungal agents to kill the cells. The production of persister cells is done by shutting down of energy conserve processes and upregulation of genes involved in growth arrest.[6] Priority of the cell is not to enter into dormancy rather remain as dividing cell as it takes more time for dormant cell to come back to original normal state.[5] The best way to achieve persistence is the upregulation of genes involved in toxin production. Overproduction of toxic products inhibits important cellular processes and cell growth, which ultimately inhibits the antibiotic action as the target for antibiotic or antifungal agents, is not expressed by cell.[7] The drug resistance mechanisms include modification of target by mutation, target modification by enzymes or producing alternative targets. The main objective of these modifications is to prevent antibiotic binding to target. If persister cells are dormant and not actively growing, no cell-wall synthesis, very little translation then antibiotics will act on these cells but due to absence of target, antibiotics will not be able to change or corrupt the target molecule. Recently, pumping persister were discovered that uses efflux pumps and throw the antibiotic out from the cells to keep the intracellular concertation of

antibiotics at lower level.[8] Another study reported found that depletion of ATP plays an important role in persister formation in *S. aureus*.[9] Microbes achieve their persistence by above mechanism and become tolerant to multidrug but they also pay a cost as being remained non-dividing or by slow growth.[6] It is known that most of chronic infections are due to persister cells but we do not have a single proficient drug to eradicate these persister cells. Many of antibiotics would fail if tested against non-growing stationary phase but all will collapse if tested against dormant cells. The first to handle chronic infections is to develop antibiotics and antifungal therapies against non-growing, dormant/persister cells that will not be an easy step.[7]

Biofilms and Their Role in Persister Cells Formation

Numerous microorganisms have the capacity to grow surfaces in a collection of cells called a biofilm. They became very important topic in research because the recognition of their role in different infections in humans.[10] Biofilms are blamed for a wide range of microbial infections in human as cells present in biofilms are more resistant or tolerant to drugs than free living cells.[10] Fungal biofilms play an important role in highly unmanageable infections. The formation of persister cells in bacteria and fungi is different such as yeast persister cells were only found in biofilm rather than in a non-growing stationary phase population as it happens in bacteria.[7] The persister cell formation in fungal biofilms is quite different bacterial persister cells in which planktonic or free living cells contains higher number of persister cells than biofilm.[6] This observation showed that biofilm is the mode of survival or production of persister cells for yeast and is the main source of formation of drug tolerant cells. The production of drug tolerant persister cells is an important way of giving cell heterogeneity to yeast population.[11]

Introduction to Cell Encapsulation Methods

Yeast cells are decent shielding structure. These cells have been used as encapsulation materials for different applications for last 40 years. They have been used in numerous fields such as pharmaceutical, food processing, and environmental protection for many years. However, these cells need favorable growth conditions to work efficiently. In order to provide optimum conditions for growth to microbial cell, cell immobilization techniques have been proposed.[12] Cell immobilization is imprisonment of whole cells to a confined space while keeping cell viable.[13] Numerous methods have been developed for cell immobilization such as adsorption or attachment of cells to an inert substance, self-aggregation by flocculation and encapsulation or confinement

using polymers.[14] To ensure cells viability, confinement is done under moderate conditions.[15]

Extrusion

In this method, microbial cells are mixed with a polymeric solution, which are later extruded through an opening as droplets into the solution of a cross-linking agent. The mixing of polymer solution with cross-linking agent results in solidification or gelation. There are number of factors affect the size of spheres include the diameter of opening, the thickness and movement, concentration and temperature of polymeric solution.[16] Extrusion is simple, inexpensive, and moderate operation conditions that ensure viability of cells.[15]

Encapsulation in Silica Particles via Spray Drying

The Spray Drying method uses the prehydrolyzed silica sol plus short chain lipid solution but try to make sure minimal contact between cells and somewhat cytotoxic precursor sol constituents by mixing the cells and sol immediately prior to introduction to the spray nozzle. Cells and precursor are combined via an automatic feed (peristaltic pumps) and aspirated into a heated, dry N2 gas sheath forming small liquid droplets. Droplets dry very rapidly (within ~400 ms), forming ordered lipid/silica mesophases in a manner related to aerosol-assisted EISA, yielding solidified particles with size distributions from ~0.5 to 25 μm.[17] There is need to optimize the conditions such as air movement, temperature, inlet and outlet air temperature.[14]

Encapsulation in Silica Films via Cell Directed Assembly

The integration of cells into physical devices is very challenging. Scientist believed that there is need to replace dishes with to 3D-dimensional designs that resembles closer to cell's natural extracellular matrix. In this method, cells are mixed with a soluble form of silica plus a short chain phospholipid (DiC6PC) plus water and ethanol and HCl. These samples are rapidly deposited on a substrate by spin-coating technique wherein evaporation of the ethanol and water drives the self-assembly of the lipids into a periodic mesophase. Water interface is replaced with silica. So the cells are incorporated in a hexagonal or lamellar lipid-silica mesophase.[18] The nanostructure used in this method of encapsulation is made of very small (nanometer-scale) channels with a thin size delivery that avoids widespread drying and associated stresses. Cell directed assembly maintains three-dimensional fluidic connectivity and allows simplistic cellular assimilation into devices.[18]

Aqueous Sol-Gel Process for Cell Encapsulation

This method employs an aqueous sodium silicate precursor solution instead of the silicon alkoxide-based precursor solution as used in above two methods. The sodium silicate solution is hydrolyzed by ion exchange to create an acidified, aqueous silicic acid sol (pH 3), which upon addition of cells in buffer, forms a porous silica gel monolith promoted by the accelerated rate of silica condensation at neutral pH. Glycerol is added to the sol precursor solution prior to gelation and serves as an ameliorant against stresses exerted on the cells stemming from syneresis. The precursor solution and cells in buffer (as before) are mixed briefly and immediately dispensed into a vessel. The solution soldifies into a gel within ~15-30s and is further aged for 12-24hrs to promote more complete silica condensation. The gel consists of a highly porous network of silica nanoparticles, which physically entraps cells plus buffer and glycerol.[19]

Platforms to Measure Gene Expression

Microarray experiments are considered as a major platform to measure the gene expression that allows to quantify the average mRNA expression level. A DNA microarray is a collection of microscopic DNA spots attached to a solid surface that is used to detect the presence and abundance of labeled nucleic acids in samples on a high throughput level. Each DNA spot in microarray contains specific DNA sequence known as probes. Probe is a small portion of a gene or other DNA element that is supposes to hybridize a target, which can be cDNA or cRNA under high rigor conditions. To identify which genes are expressed under given condition in experiment, amounts of mRNA produced by cells are measured or examined. Differential expression of mRNA in turn uncovers how cells adjust its internal homeostasis in response to outer environment. There are two types of DNA microarrays: 1) oligonucleotide arrays, 2) a variety of cDNA arrays. Affymetrix *in-situ* synthesized oligonucleotide array is the most commonly used microarray platform to measure the differential gene expression.

Affymetrix in-situ synthesized oligonucleotide

Oligonucleotide probes used in affymetrix are a set of 20-25 short oligonucleotides. The probes are specific for each gene or exon. Furthermore, one base mismatch is integrated at middle position of each oligonucleotide. Probes are made in situ using genomic information to lead photolithographic deposition. The arrays are hybridized to a single biotinylated amplified RNA

sample. Gene expression is measured for each gene taking the difference between the match and mismatch measurements, and averages. In affymetrix, image-processing algorithms are included into the GeneChip experimental process as well. Oligonucleotide arrays are powerful tools for monitoring gene expression in comparison to cDNA microarrays as later one comprises large pieces of DNA that can be from hundreds to thousands of base pairs. There are some other advantages to oligonucleotide-based microarrays over cDNA arrays such as the quality and reproducibility of printed oligonucleotide arrays is excellent.[20] In our study, we used Yeast Genome 2.0 Array that contains "5,744 probe sets for 5,841 of the 5,845 genes" present in *S. cerevisiae*.

Microarray Data Processing and Normalization

DNA microarray technology is very strong way for genome wide research. DNA microarray results in measurement of differential gene expression of hundreds of thousands of genes across different conditions. Due to limitation of technology, there are some differences in RNA quantities and variations such as intensity levels may differ from one replicate to another replicate. There is need to take some important steps before analysis of microarray data for differential expression of genes. Due to these differences, normalization is used to remove systematic errors and bias originated from microarray experimentation. Normalization methods try to remove sources of variation helps in meaningful biological analysis.[21] As it was described earlier that affymetrix array is widely used method for measurement of gene expression data. Affymetrix technology comprise perfect match and mismatch probe pairs along with probes for each individual gene. To measure the absolute expression for each specific gene, meaningful data processing is required. In preprocessing steps, multiple probe signals are combined into one absolute call, which is also referred as normalization methods. Therefore, it is very vital step to choose appropriate normalization method as normalization has an important effect on selection of differential genes. There are different normalization methods as briefly described below.

Microarray Suite 5.0 (MAS5)

Affymetrix has developed Microarray Suite 5.0 (MAS5) normalization method. MAS5 first correct both perfect match (PM) and mismatch (MM). In second step, mismatches are converted into ideal mismatches in such way that perfect match values greater than their corresponding mismatch as 30% of time mismatch values are greater than their respective perfect match. In another case, if

mismatch value is already lesser than the respective perfect match then mismatch value remain unchanged. In next step, algorithm mean is calculated over log2-transformed differences between perfect match and mismatch followed by normalization of intensity values by adjusting the timed mean of original signals to an already specified value. MAS5 normalize the data after the summarization step same as many other algorithms do.[22]

*Robust Multi-Array Analysis (RMA***)**

Robust Multi-array algorithm convert probe level data into gene expression values. This normalization method ignores mismatch probe reading. However, MM probes provides very useful information but according to RMA developers, mismatch probes incorporate more noise. RMA adjust background noise on a raw intensity scale such that there is no negative background corrected values. Furthermore, the log2-converted values are obtained for each background corrected perfect match probe followed by quantile normalization.[23]

GeneChip RMA (GCRMA)

GeneChip RMA (GCRMA) is commonly used technique for microarray normalization. GCRMA is mainly based on RMA normalization method. The difference between GCRMA and RMA is that the former one considers probe sequence information to measure the background. GeneChip RMA (GCRMA) use quantile normalization to process the gene expression data. GeneChip RMA (GCRMA) algorithm considers bases composition and location on the probe to adjust the intensity values as hybridization between complimentary probe and sample bases largely depend on labeling such that label usually bind to Cytosine that can affect the cytosine and guanine bond. To account these factors, GCRMA included the adjustment formula for normalization of gene expression data. GeneChip RMA (GCRMA) is advance form of RMA normalization method and it is more accurate method than other normalization methods.[22]

Introduction to Functional Analysis

Microarray technologies have been widely used for gene expression analysis and huge amount of gene expression data and knowledge has been produced. There is need of robust computational methods for incorporation, extraction, comparative analysis, and biological meaning of high-throughput data. There are numerous bioinformatics methods to thoroughly isolate large gene lists

to build a summary of the most enriched and related biology. The main objective of functional analysis is to find different classes differentially enriched among genes.[24] To analyze large set of genes, Gene Set Enrichment Analysis (GSEA) can be performed that uses all data on microarray rather than selected list of differentially expressed genes based on some threshold.[25]

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) Knowledgebase

The DAVID resources consist of an integrated biological knowledgebase and analytic tools designed at thoroughly mining biological meaning from large gene/protein lists.[26] The DAVID database provides a full set of tools/algorithms for understanding biological functions associated with large set of genes. The DAVID database contains cluster of information for genes and their products from different public databases. Numerous tools in DAVID database are efficient in finding enriched Gene Ontologies (GO), determining enriched functional-related gene groups; visualizing genes on BioCarta[27] and KEGG[28] pathway maps and converting gene identifiers from one type to another. The DAVID database use modified Fisher's exact test and EASE score that is a modified Fisher Exact p-value , which is 0.1 by default.[26]

Gene Ontology (GO)

Gene ontology (GO) project is a major bioinformatics ingenuity to merge the representation of gene and gene product attributes across all species. The main aims of GO project is 1) Provide and develop the organized vocabulary of gene and their products, 2) Interpret the functions of genes their products, and integrate and distribute annotation data, 3) Offer tools to extract features of the data, and 4) Allow functional biological understanding of data using the GO through enrichment analysis.[29] There are following three main classes for Gene Ontologies (GO): 1) Biological Process, are series of events with defined start and end, crucial to cell's function, 2) Molecular Function, the fundamental actions of a gene product at the molecular level, and 3) Cellular component, is the location in cell where gene's products are located or functioning.[30]

Gene Set Enrichment Analysis (GSEA)

Gene set enrichment analysis is a method to select the group of genes overrepresented or underrepresented in large list of genes and show the association of these genes with known diseases or phenotypes.[25] In GSEA, all genes (for example all genes on chip) are used in analysis rather

than selected differentially expressed genes. In this method, all genes on chip are ranked listed (L) according to some measure such as fold change. GSEA uses prior gene sets who have been associated with known phenotype. These predefined gene set (S) is mapped to ranked list (L) genes to find the location of each gene from genes set S in ranked list (L) genes. The enrichment score is calculated by walking down in the list L and if gene in S mapped in list L then running sum statistics increased and running sum statistics decreased when a gene in L is not present in gene set S. The significance of enrichment score (ES) is calculated by permutation test procedure, i.e. permute the phenotype labels or random gene sets and recomputed the ES.[25]

Multiple Hypothesis Testing

Multiple hypothesis testing refers to the testing of more than one hypothesis simultaneously. For example, t-test used on number of genes and those genes whose p-value is lower than the arbitrary selected threshold such as 0.05 can be selected as differentially expressed genes. As microarray or next generation sequencing studies generate huge amount of data such as on scale of 10000 genes and using 0.05 *p*-value there will 500 genes who has *p*-value lower than 0.05 by chance and will considered as false positives and False discovery rate (FDR) control is one of the methods used to correct multiple tests. In 1995, Benjamini Hochberg proposed a method called FDR Benjamini Hochberg method to correct multiple comparisons and it is most widely used method for controlling false positives.[31] In this method, p-values are sorted and then ranked those p-values. The smallest p-value get the rank 1 and largest will get the rank N. In next step, each p-value is multiplied by N, which followed by division by its assigned rank that results the adjusted p-value. Bonferroni method is also used to correct multiple hypothesis and it is considered as very conservative method.[32] In Bonferroni method, each p-value is multiplied by the total number of genes (N) tested. The resulted p-value will be very high and only few genes will have selected as differentially expressed.

Permutation Tests

Permutation tests are statistical significance tests that do not assume any distribution to estimate the significance of test statistic. Permutation test is also called randomization test or exact permutation tests and can be used for any test statistics. Exact permutation test use all possible values with replacement of test statistics and it turns out be computationally very expensive.

However, in case of randomized permutation, only subsets of test statistics are used to calculate or approximate the sampling distribution rather than using all possible values of test statistics. In random permutation tests, significance level or confidence level can be increased by increasing the number of random samples.[33]

Gene Network Visualization and Analysis

Visualization and analysis of data generated from high-throughput techniques is an important technique to know the comprehensive relationship of genes and proteins. In networks, genes or proteins are represented as nodes and edges represent the relationship between these genes or proteins. The interaction between network entities can be direct or indirect. Cytoscape[34] is widespread open source tool for visualization of gene network interaction using gene expression profiles or other types of measures. BisoGenet is one the plugin used in cytoscape software for building and visualization of gene network.[35] BisoGenet consist of three levels such as data level, middle level and client level. The data component stores protein-protein interaction (PPI), gene-DNA interaction, Gene Ontology (GO) and pathway information. In middle component, the bioentities and their relationship from database is represented as a global network. The third level is called client tier that is a Cytoscape plugin, which takes and manage user input followed by communication with web service and finally visualize and analyze the network. String (Search Tool for the Retrieval of Interacting Genes) is a database of known and predicted protein-protein interaction that provide physical as well as functional association.[36] It contains biological information from various sources such as experimental data, computational data, data mining, co-expression data. The latest version of string contains 9.6 million proteins from about 200 organisms. String also provides functional enrichment of input list of proteins on carious domains such GO and pathway.

Chapter 2: Three-Dimensional Encapsulation of Saccharomyces Cerevisiae in Silicate Matrices Creates Distinct Metabolic States as Revealed by Gene Chip[1]

Abstract

To design hybrid cellular/synthetic devices such as sensors and vaccines, it is important to understand how the metabolic state of living cells changes upon physical confinement within three-dimensional matrices. We analyze the gene expression patterns of stationary phase *Saccharomyces cerevisiae* (*S. cerevisiae*) cells encapsulated within three distinct nanostructured silica matrices and relate those patterns to known naturally occurring metabolic states. Silica encapsulation methods employed were lipid-templated mesophase silica thin films formed by cell-directed assembly (CDA), lipid-templated mesophase silica particles formed by spray drying (SD), and glycerol doped silica gel monoliths prepared from an aqueous silicate (AqS+g) precursor solution. It was found that the cells for all three-encapsulated methods enter quiescent states characteristic of response to stress, albeit to different degrees and with differences in detail. By the measure of enrichment of stress-related Gene Ontology categories, we find that the AqS+g encapsulation more amenable to the cells than CDA and SD encapsulation. We hypothesize that this differential response in the AqS+g encapsulation is related to four properties of the encapsulating gel: 1) oxygen permeability, 2) relative softness of the material, 3) development of a protective sheath around individual cells (visible in TEM micrographs *vide infra*), and 4) the presence of glycerol in the gel, which has been previously noted to serve as a protectant for encapsulated cells and can serve as the sole carbon source for *S. cerevisiae* under aerobic conditions. This work represents a combination of experiment and analysis aimed at the design and development of 3D encapsulation procedures to induce, and perhaps control, well-defined physiological behaviors.

Introduction

Harnessing unique properties innate to biomolecules and living cells by incorporating them within

[1] Reprinted with permission from (**Three-Dimensional Encapsulation of Saccharomyces cerevisiae in Silicate Matrices Creates Distinct Metabolic States as Revealed by Gene Chip Analysis***: Zeeshan Fazal, Jennifer Pelowitz, Patrick E. Johnson, Jason C. Harper, C. Jeffrey Brinker, and Eric Jakobsson. ACS Nano 2017 11 (4), 3560-3575 DOI: 10.1021/acsnano.6b06385).* Copyright (2017) American Chemical Society.

hybrid materials continues to excite and inspire the efforts of researchers.[37–40] Such functional hybrid living materials may impact areas as diverse as biocatalysis, controlled delivery of therapeutics, stabilization of probiotics in food products, environmental and human health monitoring, industrial process monitoring, pollution remediation, early warning of warfare agents, advanced prosthetics, bioelectronics, and tissue/organ replacement.[41–48] Extensive techniques have been developed for incorporating biomolecules and living cells within materials that can maintain biological function, while providing stability outside the *in vivo* environment. For living cells, the majority of reported techniques rely on cellular encapsulation within polymeric materials that physically confine and protect the cells and can act as a synthetic extracellular matrix.[49,50] Successful matrices incorporate specific material properties that assuage chemical and mechanical stresses exerted on the integrated cells, provide access to oxygen, nutrients and molecules of interest, and allow for expulsion of metabolic wastes.[51–54] Many reports have implicated the importance of the interface chemistry between the cell surface and host matrix on meeting these requirements and enhancing long-term viability and activity.[55,56]

Encapsulation and physical confinement of living cells, however, can drastically alter the cellular environment and exert multiple stresses on the cells. Stresses commonly associated with encapsulation include compressive/tensile stress, osmotic stress, oxidative stress, acid/base exposure, nutrient depletion, temperature shock, and cytotoxic chemical exposure.[57–59] Unicellular organisms have evolved to rapidly respond to abrupt changes in environmental conditions by autonomous mechanisms that regulate gene expression profiles.[3] By altering the cell's metabolic state, cells can often maintain the vital functions and processes necessary for viability and growth.[60,61] As an example, some microorganisms encapsulate themselves in polymer matrix biofilms. These cells sense changes in their local environment (*e.g*. quorum sensing) and alter their metabolic state, showing enhanced resistance to antibiotics, upregulation of virulence factors, development of competence, and changed growth morphology.[62]

As an alternative to organic polymers, amorphous silicon dioxide (SiO_2) and hybrid organosilicate matrices have also been used for cellular encapsulation. Generally, silica matrix encapsulated cells are physically confined such that growth and division is arrested. Metabolites and waste products must diffuse through the silica matrix, and the cell wall/membrane may interact with the polar and

negatively charged silica surface, which is terminated with surface silanol (≡Si-OH) and deprotonated silanol (≡Si-O$^-$) groups. The metabolic state of living cells has been shown to change significantly upon physical confinement within silica matrices. For example, in response to this 3D encapsulation, *Catharantus roseus* plant cells have been shown to substantially increase secondary metabolite (*i.e*. alkaloid) production, demonstrating a 10-100 fold increase over cells free in solution.[63] This may be due to the cells shifting metabolic load from growth/division towards other metabolic pathways. Sol-gel immobilized *Escherichia coli* (*E. coli*) cells have also been reported to express GFP,[64] and produce ATP,[65] at levels nearly 2-fold greater than cells in solution. In these studies, the authors did not provide a hypothesis for this significant enhancement in biosynthesis. An individual *Staphylococcus aureus (S. aureus)* cell encapsulated in a nanostructured silica matrix was observed to self-initiate quorum sensing (QS) pathways due to constrained diffusion of quorum sensing 'autoinducer' signaling molecules, showing that confinement alone can serve as a mechanism for inducing QS and its associated metabolic shifts.[66] Further, encapsulated QS *S. aureus* cells[66] and encapsulated QS *Serratia marcescens (S. marcescens)* cells[67] have significantly longer viability over encapsulated non-QS *S. aureus* and *S. marcescens*, indicating profound effects of encapsulation that have not as yet been well documented. The metabolic state of cells upon encapsulation can also significantly impact their ability to respond to the stresses of encapsulation. *Saccharomyces cerevisiae* cells from exponential phase culture (high metabolic activity) showed higher rates of induced gene expression than cells encapsulated from stationary phase cultures (low metabolic activity).[68] However, cells from stationary phase cultures exhibited significantly greater long-term viability. Further, addition of nutrients to the encapsulation matrix, which researchers expected to enhance long-term viability,[55] adversely impacted long-term viability. This was attributed to media components inducing exit of the cells from more robust metabolic states (*i.e*. quiescence), and the metabolic production of toxic byproducts.

Despite significant efforts by many researchers to elucidate and characterize the condition of cells upon encapsulation, the physiological and metabolic state of these cells remains poorly understood. This is in part due to the challenges associated with probing cells within the encapsulation matrix and measuring the cellular response. Specifically, the 'gold-standard' measurement of viability is reproductive capability. As the vast majority of silica matrices

physically restrict cells from growth and division,[56] (with notable recent exceptions[64,69]) this technique for evaluating viability requires arduous procedures to free the cells from the matrix without compromising their integrity. Removal from the matrix may again alter the cellular state, or encapsulated cells may have entered an irreversible viable but non-culturable (VBNC) state[17,70] such that they maintain metabolic activity by do not reproduce, making it challenging to draw conclusions from these measurements.[65]

To avoid these complications, it has become increasingly common for researchers to employ vital dye staining and fluorescence microscopy to report viability of encapsulated cells.[71] Vital dye staining typically assays cell membrane integrity and non-specific enzyme activity; however, this technique has been shown to not always correlate with reproductive capability. Davey and Hexley reported that *S. cerevisiae* subjected to chemical and physical stresses stained with propidium iodide (PI), a membrane impermeant dye, indicative of cell death.[72] However, they also showed that some of the PI stained cells were able to recover and reproduce. Further, we recently reported that *E. coli* cells encapsulated in an aqueous silicate-based matrix showed compromised membrane integrity, but enhanced gene expression rates over cells in solution and retained reproductive capability when freed from the matrix.[73] This lack of correlation between membrane integrity, reproductive capacity, and protein expression highlight the complex relationship between the nanostructured inorganic matrix chemistry and the metabolic state of the encapsulated biological. Herein we present the genomic expression patterns of *S. cerevisiae* cells encapsulated within silica matrices and relate those patterns to known naturally occurring metabolic states. Within the field of 3D cellular encapsulation, this is the first paper to comprehensively elucidate how systematic modifications of the encapsulating nanostructure influence cellular behavior based on gene chip analysis. In this work, three distinct silica encapsulation methods were employed: lipid-templated mesophase silica thin films formed by cell-directed assembly (CDA),[18] lipid-templated mesophase silica particles formed by spray drying (SD),[17] and glycerol doped silica gel monoliths prepared from an aqueous silicate (AqS+g) precursor solution.[33] *S. cerevisiae* cells from stationary phase cultures were used for encapsulation and as the reference condition for identifying differentially expressed genes. Gene enrichment analysis and grouping of biologically significant attributes *via* gene ontology (GO) revealed that distinct metabolic states were induced in *S. cerevisiae* from each encapsulation method. Further, the material properties at the bio-nano interface more profoundly

influence biological behavior than the bulk chemistry of the matrix and the effect of physical confinement alone. Additionally, analysis of enriched gene data revealed the significant unique stresses exerted on cells from a encapsulation method and the cellular response to those stresses. This information will allow for informed tuning of encapsulation matrix parameters to improve upon desired biological traits, and may facilitate the ability to induce, and perhaps control, new and unique physiological behaviors.

Materials and Methods

Materials

Saccharomyces cerevisiae (strain S288c) was acquired from American Type Culture Collection (ATCC; Manassas, VA). 1,2-dihexanoyl-sn-glycero-3-phosphocholine (*di*C_6 PC) was purchased from Avanti Polar Lipids, Inc. (Alabaster, AL). Absolute ethanol, hydrochloric acid (HCl, 37%), tetraethylorthosilicate (TEOS, 98%), sodium silicate solution (26.5% SiO_2; 10.6% Na_2O), strongly acid cation-exchange resin (DOWEX 50WX8-100, hydrogen, 50-100 mesh), glycerol, yeast extract, peptone, D-(+)-glucose, adenine, sodium acetate, ethylenediaminetetraacetic acid (EDTA), sodium dodecyl sulfate (SDS), phenol, choloroform:isoamyl alcohol (24:1), sodium phosphate (mono- and dibasic), sodium acetate, phosphate buffer saline (PBS, pH 7.0), glass beads, and Alconox detergent were purchased from Sigma-Aldrich (St. Louis, MO). RNeasy Mini Kit (including Buffer RW1 and Buffer RPE) was obtained from Qiagen (Germantown, MD). RNA 6000 Nano Kit was purchased from Agilent (Santa Clara, CA). Yeast Genome 2.0 Arrays and Genechip 3' IVT Express Kits were from Affymetrix (Santa Clara, CA).

Cell Culture

S. cerevisiae S288c cells were grown aerobically in Yeast Peptone Dextrose + Adenine (YPD+A) broth at 30 °C for seven days until the culture was in stationary phase (OD600 1.5-2.0). Cells were then centrifuged at 10,000 rpm for 5 mins, washed with DI H_2O three times, and re-suspended in DI H_2O (OD600 0.1-1.0).

Encapsulation of S. cerevisiae in Phospholipid-Templated Silica Films via Cell Directed Assembly (CDA)[18]

Glass coverslips (25-mm, No. 1.5) were soaked in 0.1 M KOH for a minimum of 4hr, washed with

10% Alconox (m/v), rinsed with DI H_2O, dried under a stream of N_2 gas, and then cleaned in UV-ozone (Jelight, model 342) for 5mins. Prehydrolyzed tetraethyl orthosilicate (TEOS) stock solutions (A2**) were made by refluxing 61 mL of TEOS, 61 mL of absolute ethanol, 4.9 mL of DI H_2O, and 0.2 mL of 0.07 N HCl (molar ratio $1:4:1:5x10^{-5}$) for 90mins at 60°C. A2** stock solutions were then stored at -20°C. CDA precursor sol was prepared by combining 0.25 mL of A2** stock solution, 0.2 mL of absolute ethanol, 0.16 mL of 0.05 N HCl, 0.4 mL of DI H_2O and 30 mg of *di*C_6 PC. This precursor solution was aged at room temperature for 20mins and then passed through a 0.45-μm filter. To this solution was combined an equal volume of PBS washed and re-suspended *S. cerevisiae*. This mixture was cast onto cleaned glass coverslips. Resulting thin films could air dry, were rinsed with DI H_2O to remove non-integrated cells, and stored under ambient conditions (25˚C ± 2˚C) for three days.

Encapsulation of S. cerevisiae in Phospholipid-Templated Silica Particles via Spray Drying (SD)[17]
The silica precursor sol was prepared by adding 0.83 mL A2** stock (described above) to a solution containing 1.33 mL DI H_2O, 0.66 mL ethanol and 0.53 mL 0.07 N HCl – the same chemistry as the CDA method, but scaled-up for higher material yield. This solution could age at room temperature with sonication for 30-60mins until complete homogenization of the precursors had occurred. Immediately prior to sample preparation, 100 mg diC6 PC lipid was added to the fully condensed sol and mixed until fully dissolved (~20s).

Samples were spray dried with a Mini Spray Drier B-290 (Buchi, Flawil, Switzerland) using a 0.7 mm nozzle. The inlet temperature of the spray nozzle was adjusted between 60-65˚C to maintain an outlet temperature of 30˚C, with an aspiration rate of 90%, a peristaltic pump feed rate of 3.5 mL/min, and nitrogen as a carrier gas at a rate of 60 L/hr. 3.3mL of precursor sol and dissolved lipid (described above), and 3.3mL of cells in PBS suspension, were loaded into separate assimilation vials. Two peristaltic pumps with individual feed rates of 1.75 mL/min (3.5mL/min combined feed rate) were used to deliver the solution to the spray drier nozzle with mixing *via* a Y connector immediately prior to inspiration into the nozzle. This technique allowed for minimal contact between cells and the acidic and alcoholic precursor solution, improving cell viability. After mixing, the solution containing cells, silica precursors and lipid was aerosolized into the dry nitrogen sheath established within the spray drier. Rapid evaporation of the solvent results in lipid-

directed evaporation induced self-assembly (EISA). Particles are fully dried before entering the spray drier cyclone. Spray dried particles were collected in assimilation vials that were connected to the standard cyclone, replacing the standard collection chamber. After spray drying, samples were stored under ambient conditions (25˚C ± 2˚C and 60% RH ± 5% RH) for three days.

Encapsulation of S. cerevisiae in Glycerol Doped Aqueous Silica Gel Monoliths (AqS+g)[73]

An aqueous silica matrix precursor solution was prepared by adding 1.654 mL of sodium silicate solution to 6.8 mL of DI H_2O. This mixture was immediately added to 3.08 g of highly acidic hydrogen (H^+) cation-exchange resin. The solution was mixed for approximately 2mins. The resin was then removed by vacuum filtration, or by centrifugation to pellet the resin, followed by collection of the supernatant. To this solution, 20% (vol.) glycerol was added. Stationary phase *S. cerevisiae* cells were washed 3× and resuspended in 1.0 M sodium phosphate buffer, pH 7.0. Cells in buffer and aqueous sol precursor solution were then mixed in a 1:1 ratio. Gelation occurred within 25-30 seconds of mixing the two solutions. Final cell density was 10^6-10^7 cells/mL. Silica gels were formed and stored under ambient conditions (25˚C ± 2˚C) for three days.

Optical Microscopy

For optical imaging, dried powders were suspended in water, vortexed for 10s and pipetted onto standard microscope slides. Samples were imaged on a Zeiss LSM 510 confocal microscope mounted on a Zeiss Axiovert 100 inverted microscope. Prior to encapsulation, *S. cerevisiae* were stained with Syto-9 green fluorescent dye according to manufacturer's specifications.

We created z-stack images for particles of varying sizes to visualize the distribution of cells within particles. This was achieved by setting the upper and lower boundaries of a particle and taking an image with a given optical slice diameter and collecting an image every diameter distance. The resulting collection of images maps the entire z-dimension within the sample, allowing us to create 3D reconstructions of the sample. Compressed z-stack images were created by merging all the images from the z-stack into one image, allowing for easier visualization of the entire particle.

Electron Microscopy

Prior to transmission electron microscopy (TEM), image contrast between *S. cerevisiae* cells and the encapsulating matrix was improved for all three matrix types *via* fixation and staining of cells

using glutaraldehyde and osmium tetroxide respectively. For CDA and SD samples, this was accomplished simply by adding the powder to the solutions outlined below. For the gel sample, a cured gel (300µL volume, *vida infra*) was fractured into smaller ~1mm pieces using a clean razor blade prior to treatment. We fixed the cells by incubating the samples in 2.5% glutaraldehyde in PBS for 1hr at RT (room temperature) followed by 3x rinse cycles in PBS using 10kRPM and 30s for centrifuge settings. We stained the samples using osmium tetroxide (OsO_4) following the protocol recommended by R. Wright.[74] Here, fixed cells are centrifuged, the supernatant is removed, and 1% OsO_4 in PBS is added and the sample is incubated for 2hrs on ice due to the high vapor pressure of OsO_4.

Prior to TEM imaging, it was necessary to remove the aqueous phase within the samples, which was done by stepwise dehydration and water replacement with ethanol. This was achieved by incubating the three fixed and stained samples in increasing concentrations of EtOH for 15mins (concentrations were 30%, 50%, 70%, 90%) and finally for 60mins (concentrations were 95% and 100%). Dehydrated samples were stored at 4°C. For TEM imaging, samples were switched to an anhydrous acetone for the final dehydration. The preparation was then infiltrated with resin by incubating particles in 1:1 Spurr's resin: acetone, 3:1 Spurr's resin: acetone and, finally, 100% Spurr's resin. Samples were placed in embedding molds, polymerized by incubation at 60˚C for at least 16hrs, and the blocks were trimmed for microtoming. Microtomed sections with thicknesses between 60 and 80nm were used for imaging. Transmission Electron Microscopy was performed using a Hitachi H7700 equipped with an AMT XR16M 16-megapixel camera.

Extraction and Purification of RNA from Encapsulated Cells

Portions of encapsulated cell-containing silica matrices were rinsed with DI H_2O and placed in 2 mL centrifuge tubes, breaking the matrix into smaller pieces as necessary. 270 µL sodium acetate buffer (50 mM sodium acetate, pH 5.2, and 10 mM EDTA, pH 8.0), 30 µL 10% SDS, and approximately 100 µL of glass beads (0.5 – 0.6 mm diameter) were added to centrifuge tubes. Samples were stirred and agitated to physically break apart the matrix. Water saturated acid phenol (300 µL) preheated to 65°C was added to the samples, followed by vortexing for 1min. Samples were then incubated in a 65°C water bath for 5mins, followed again by vortexing for 1min. This was repeated for a total of six cycles over approximately 45mins. Samples were then placed in a

cooling block for 2mins, followed by the addition of 300 µL choloroform: isoamyl alcohol (24:1), and mixed by vortexing for 30 seconds. Samples were then centrifuged for 5mins at 15,000 rpm, followed by transferring the RNA-containing top layer to new centrifuge tubes. An equal volume of 70% ethanol was added and mixed by pipetting.

Next, samples were transferred to Qiagen RNeasy spin columns and centrifuged for 30 seconds at 15,000 RPM, discarding flow through. 700 µl of RW1 buffer was added to the columns and the samples were centrifuged for 30 seconds at 15,000 RPM, discarding flow through. 500 µL of RPE buffer was added to the columns, followed by centrifugation for 30 seconds at 15,000 RPM, discarding flow through. 500 µL of RPE buffer was added to the columns and followed by centrifuging for 2mins at 15,000 RPM, discarding flow through. Columns were then placed in new 2 mL collection tubes and centrifuged for 1min at 15,000 RPM. Next, columns were placed in new 1.5 mL collection tubes and 40 µL of DI H_2O was added. Columns were again centrifuged for 1min at 15,000 RPM, after which 40 µL of DI H_2O was added and the columns were centrifuged again for 1min, yielding the purified RNA product. A NanoDrop spectrophotometer (Wilmington, DE) and an Agilent 2100 Bioanalyzer (Santa Clara, CA) were used to determine concentration, purity and integrity of RNA extractions. RNA samples were stored at -20°C until used.

Gene Chip Hybridization

cDNA preparation, and biotin-labeled cRNA generation and fragmentation were performed using Affymetix Genechip 3'IVT Express Kits per the manufacturer protocol. Hybridization solution was prepared with the fragmented cRNA and controls, and was then hybridized to the probe array with a 16-hour incubation period. The solution was then removed and the probe array was washed and stained in an automated microfluidics station. Arrays were read by an Affymetrix Genechip Scanner attached to a workstation running Affymetrix Microarray Suite.

Experimental Design

Gene chip data from four treatments groups were collected. Each treatment had two replicates. Group 1: Stationary phase *S. cerevisiae* cells (grown for 7 days in aerated media) washed and resuspended in phosphate buffer saline solution (PBS). This group served as the baseline cultured *S. cerevisiae* control to which the RNA expression levels from sample groups 2-4 were compared.

Group 2: Stationary phase *S. cerevisiae* cells encapsulation in silica *via* cell-directed assembly (CDA) and stored dry for three days under ambient temperature and relative humidity (RH). Group 3: Stationary phase *S. cerevisiae* cells encapsulation in silica *via* spray drying (SD) and stored dry for three days under ambient temperature and RH. Group 4: Stationary phase *S. cerevisiae* cells encapsulation in glycerol doped aqueous silica gels and stored sealed for three days under ambient temperature.

Identification of Differentially Expressed Genes

Statistical analyses were performed with the LIMMA[75] R package using an empirical Bayes linear modeling approach and gene expression intensities were normalized using GeneChip RMA (GCRMA)[76] normalization. To control multiple hypothesis testing, Benjamini and Hochberg (B&H)[31] FDR method was used to correct raw *p*-values. The resulting statistical analyses provide the differentially expressed genes with respect to control based on estimate measures of gene expression, with associated *p*-values. The FDR-adjusted *p*-value <0.05 and (|log2(fold change) | > 1 was used to identify the differentially upregulated and downregulated genes respectively between CDA vs Control, SD vs Control, and AqS+g vs Control.

Functional and Gene Network Analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID)[26] was used to identify enriched GO terms. This functional analysis permitted the identification of Gene Ontology (GO)[30] from the gene expression data, including biological processes, cellular component, and molecular function. GO terms were considered significantly enriched at FDR-adjusted *p*-value <0.05 using the Yeast 2.0 array as background. Redundant GO terms were removed and further summarized by semantic similarity in REVIGO,[77] and represented by two-dimensional space in scatter plots. The GO terms that semantically similar should position together in scatter plot and semantic space units have no inherent meaning. The size of the circle is relative to the frequency of the GO term in underlying GO database; however, color of circle represents the log 10 *p*-value. GO terms with FDR-adjusted *p*-value <0.05 are shown on scatter plots. The threshold used for allowed sematic similarity is “medium”. Gene Ontology analysis from DAVID was complemented with Gene Set Enrichment Analysis (GSEA)[25] (**Table 2.3, 2.4, and 2.5**). GSEA was carried out on all genes present on the chip to identify the overrepresented categories in each method of encapsulation. The gene set sizes were 1000 and 15 for maximum and minimum cutoff

respectively. The statistical significance of enriched categories was measured by using 1000 permutation which was used to calculate the FDR to control the multiple hypothesis testing. Kyoto Encyclopedia of Genes and Genomes (KEGG)[78] pathway enrichment analysis was carried out using The Database for Annotation, Visualization and Integrated Discovery (DAVID)[26] to find significant pathways overrepresented in different encapsulation methods. The KEGG pathways were considered significantly enriched at FDR-adjusted *p*-value <0.05 using Yeast 2.0 background.

Gene networks were constructed and visualized using Cytoscape[34] plugin BisoGenet.[35] Networks include genes that are differentially expressed (FDR-adjusted P-value < 0.05, and (|log2(fold change) | > 5 in each encapsulation method (Supplementary Figure 2.3). Networks represent gene relationships based on protein-protein interactions annotated in BIOGRID, DIP, HPRD, INTACT, MINT, and BIND databases.

Results and Discussion

We have examined the effects of three differing silica matrices with respect to the physiological and metabolic state of encapsulated *S. cerevisiae* cells and evaluated these effects with respect to the micro- to nano- morphologies of the material structure. The methods investigated include lipid-templated mesophase silica thin films formed by cell-directed assembly (CDA), lipid-templated mesophase silica particles formed by spray drying (SD), and glycerol doped silica gel monoliths prepared from an aqueous silicate (AqS+g) precursor solution. The synthesis procedure used for each silica encapsulation method is shown schematically in **Figure 2.1** and further detailed in the Materials and Method Section.

In CDA (**Figure 2.1A**), short-chain phospholipids are used to direct the formation of an ordered lipid/silica mesophase during evaporation-induced self-assembly (EISA) of a silicate sol derived from acid-catalyzed hydrolysis of TEOS in an ethanol/water solvent.[18]

When performed in the presence of living cells, the cells actively intervene to direct the formation of a novel bio-nano interface comprising a fluid, multilayered lipid/silica mesophase that interfaces coherently with the 3D silica nanostructure.[18] To prepare samples *via* CDA, cells in buffer and lipid dissolved in the silicate sol are prepared as described in the Materials and Methods section.

These solutions are combined in a 1:1 ratio (v/v), mixed, and immediately applied to a cleaned glass slide. The solvent can evaporate (2+hrs), yielding a thick glassy coating on the slide which can be scraped into a powder mechanically using a cleaned razor blade. The resulting powder is composed of large, flake-like particles that contain many cells packed together within a nanostructured lipid/silica matrix. TEM observations (**Figure 2.1A**, material nanostructure panel) reveal an ordered, conformal lipid-silica matrix that interfaces coherently with the cell surface. Here, two cells are shown with an intervening nanostructure exhibiting a striped pattern consistent with a cylindrical or lamellar lipid/silica mesophase as noted in previous work in which silica matrix formation is ordered by amphiphilic phospholipids.[18]

The SD method (**Figure 2.1B**) employs the same prehydrolyzed silica sol plus short chain lipid solution as the CDA method, but ensures minimal contact between cells and somewhat cytotoxic precursor sol constituents (~15% ethanol v/v and pH 3) by mixing the cells and sol immediately prior to introduction to the spray nozzle. Cells and precursor solutions are prepared as with CDA samples, but they are combined *via* an automatic feed (peristaltic pumps) and aspirated into a heated, dry N_2 gas sheath forming small liquid droplets. Droplets dry very rapidly (within ~400 ms), forming ordered lipid/silica mesophases in a manner related to aerosol-assisted EISA,[79] yielding solidified particles with size distributions from ~0.5 to 25 μm.[17] As with the CDA samples, TEM images (**Figure 2.1B**, material nanostructure panel) show a conformal striped nanostructure, consistent with a lamellar or hexagonal mesophase,[18] which again interfaces directly with the cell surface.

The third encapsulation matrix is a silicate gel monolith system, AqS+g (**Figure 2.1C**), which employs an aqueous sodium silicate precursor solution instead of the silicon alkoxide-based precursor solution as used in CDA and SD. The sodium silicate solution is hydrolyzed by ion exchange to create an acidified, aqueous silicic acid sol (pH 3), which upon addition of cells in buffer, forms a porous silica gel monolith promoted by the accelerated rate of silica condensation at neutral pH.[80] Glycerol is added to the sol precursor solution prior to gelation and serves as an ameliorant against stresses exerted on the cells stemming from syneresis.[73] The precursor solution and cells in buffer (as before) are mixed briefly and immediately dispensed into a vessel (here, 1.5 mL microfuge tubes). The solution soldifies into a gel within ~15-30s and is further aged for 12-

24hrs to promote more complete silica condensation. The gel consists of a highly porous network of silica nanoparticles, which physically entraps cells plus buffer and glycerol. The silica nanoparticles are separated from the cell surface by a thin liquid layer, which may be enriched in glycerol (*vide infra*), as shown in the material nanostructure panel of **Figure 2.1C**.

Macro- and Nano-Morphologies of Encapsulated Cells *via* Cell Directed Assembly

While all *S. cerevisiae* cells in this study are encapsulated within a porous silica matrix, the physical properties at the bio-nano interface produced from each encapsulation method are markedly different. We have analyzed all three methods for their micro- and nano-morphologies using confocal fluorescence microscopy and transmission electron microscopy (TEM), and compared them with respect to their physical properties (**Figures 2.2-2.4**).

As depicted schematically in **Figure 2.1** and observed by eye, particles prepared using CDA appear rough and fragmented**. Figure 2.2A-C** are projection confocal (compressed z-stack) fluorescence and differential interference contrast (DIC) microscopy images of film fragments detailing the location of the encapsulated cells (fluorescently labeled green) within the lipid-silica fragments.

Inspection of CDA samples with TEM reveals a tightly packed array of cells incorporated within a conformal nanostructured lipid/silica matrix that interfaces directly with the cell wall (**Figure 2.2E**, inset) as reported previously.[17,18] The 'striped' nanostructure is consistent with an ordered lamellar or hexagonal lipid/silica mesophase as reported in previous work in which the formation of the silica matrix is directed by amphiphilic phospholipids.[18] Osmium tetroxide (OsO_4) staining of the cell wall allows for the visualization of the cells with respect to the lipid/silica matrix and emphasizes the conformal encapsulation of the cells. The ridge-like pattern observed in the TEM images is attributed to the high hardness (0.25 ± 0.10 GPa hardening to 0.52 ± 0.1 GPa with 15 day aging) and Young's modulus (4.3 ± 0.1 GPa, unchanging with 15 day aging) of the CDA samples, which exceed those of many natural and synthetic composites and cause "chattering" of the ultra-microtome blade as described previously.[17] This phenomenon is also evident with the SD samples, but not with the relatively soft gel materials (*vide infra*).

Macro- and Nano-Morphologies of Encapsulated Cells *via* Spray Drying.

Here, using the same precursor chemistry as with CDA, we have extended our scalable, spray drying process as developed previously for *E. coli*[17] to spray dried lipid-silica particles containing eukaryotic organisms (**Figure 2.3**). As shown schematically in **Figure 2.1**, compared to CDA, spray drying involves extremely short cell-solvent contact times due to the mixing chamber and very short residence/drying times of ~400 ms.[81] Furthermore, this technique allows for the facile preparation of large quantities of biomaterial with tight control over material macro-morphologies (size, shape, approximate cell loading), while independently maintaining the material properties (specifically nanostructure and hardness).

As observed using confocal fluorescence microscopy (**Figure 2.3 A-C**), SD particles appear as aggregates of smaller particles. Shown is an optical slice within the z-stack of the particle depicting the placement of several cells within the center of the particle.

TEM images of SD samples highlight the differing microstructure found with this class of material, which comprises individual or multiple cells incorporated within distinct, isolated macroscopic particles that preserve the conformal ordered lipid/silica nanostructure observed for CDA samples. **Figure 2.3D** highlights two *S. cerevisiae* cells encapsulated within two larger particles, showing a distinct layer of matrix "peeling" off from the surface of the cell – an artifact produced during sample preparation *via* solvent exchange or microtoming. This thin shell surrounding the cell is a further indication that the cell is completely encapsulated within the silica matrix. **Figure 2.3E** magnifies the interface between the cell and the matrix and zoomed insets clearly show an ordered lamellar structure (**Figure 2.3F**) immediately adjacent to the cell wall with a lamellar/hexagonal mesophase (**Figure 2.3G**) extending throughout the bulk of the particle. This data suggests that the encapsulation of *S. cerevisiae* cells within this class of spray dried nanostructured lipid-silica matrix exhibits essentially identical nano-properties to those observed for bacteria encapsulated in the same spray-drying procedure and exhibits a similar-sized nanostructure with ~3 nm lamellar features (**Figure 2.3F**).[17] Similar to CDA specimens, these TEM images show a distinct chattering pattern in the material microstructure, which we attribute to the high hardness (1.4 ± 0.1 GPa) and elastic modulus (13.0 ± 0.1 GPa) of the encapsulating lipid/silica mesophase. It is worth noting that, based on our previous study with *E. coli*, the reported hardness and Young's modulus of SD

samples as compared to CDA samples are ~3-6 times and ~3 times higher, respectively, and the TEM images presented herein reflect a more distinct fatigue pattern with the SD samples as compared to the CDA samples. For comparison, TEM images of gel samples (**Figure 2.4D**) exhibit no discernible chattering pattern (*vide infra*).

Macro- and Nano-Morphologies of Encapsulated Cells *via* Aqueous Gel Monoliths

Encapsulation of cells within silica gels has been extensively studied,[17] where cells are introduced into an aqueous silica sol and are physically entrapped by pH-triggered, base catalyzed silica condensation[19] with no accompanying drying, and little to no shrinkage. **Figures 2.4A-C** show confocal fluorescence and DIC images of a gel fragment, where a random incorporation of cells within a non-fluorescent silica gel fragment is observed. **Figure 2.4D** shows a TEM image of an individual cell entrapped within the silica gel matrix. Because the gel maintains the same relative volumes of the constituent components as the sol, and no evaporation occurs (as opposed to the CDA and SD samples), the cells appear individually isolated within the gel matrix (**Figure 2.4D**). This matrix is comprised of a highly porous, ramified network of aggregated 10 nm silica nanoparticles, as expected from extensive previous experiments confirming the fractal nature of amorphous silica gels.[56,82,83] Importantly, **Figure 2.4D** and the magnified inset distinctly emphasize that, compared to CDA or SD, the silica gel matrix does not conform to the cell surface and is separated by a boundary region which is rich in glycerol. Also, noteworthy as mentioned above, the gel sample is notably less hard/stiff than the CDA and SD samples, and does not exhibit any chattering during microtoming.

RNA Expression Levels of Encapsulated *S. cerevisiae* Cells

To more comprehensively explore the physiological and metabolic state induced by each silica encapsulation matrix, genetic analysis was performed using *S. cerevisiae* cells from stationary phase cultures (7 day old cultures). Cells in this state are more resistant to a multitude of stresses and have shown greater long-term viability when integrated within hybrid abiotic materials.[31-33] *S. cerevisiae* cells in stationary phase cultures have undergone two metabolic shifts. In the first shift, rapidly proliferating cells in exponential culture deplete all fermentable carbon and undergo slow carbon starvation. The cells alter their metabolism (diauxic shift) to consume ethanol and non-

fermentable carbon byproducts of fermentation. After all fermentation byproduct carbon is depleted, the cells again shift their metabolism, with a portion of the cells entering quiescence, a maintenance-like resting state where proliferation does not occur and cells can remain viable without nutrients.[84] Cells from stationary phase were used in all three silica encapsulation matrices, and were also used as the baseline condition to which encapsulated cell RNA expression levels were referenced. Identification of differentially expressed genes, and identification of gene ontology (GO),[30] followed the method of Benjamini and Hochberg,[31] by The Database for Annotation, Visualization and Integrated Discovery (DAVID), [26] as described in the Materials and Methods section.

The extent to which the three-encapsulation methods used in this work induce similar genetic expression patterns, and the extent to which they differ, is also summarized in Venn Diagram form in **Figure 2.5**. Generally, all three methods share 750 and 598 upregulated and downregulated genes, respectively. However, CDA have 427 downregulated and 301 upregulated alone, followed by SD (225 downregulated, 293 upregulated) alone. The AqS+g method has least unique number of genes differentially downregulated (138) and upregulated (124) genes.

Functional Analysis of RNA Expression

Genes for whom the biological function is known can be categorized into three GO domains: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). BP terms refer to genes associated with multistep events requiring organized assembles of molecular functions (*e.g*. RNA transport, autophagy, biosynthesis). MF terms describe genes that participate in single-step elemental activities at the molecular level, such as binding or catalysis. ATP-binding, hydrolase activity, and toll receptor binding are examples of MF terms. Genes associated with components of cells, or the extracellular environment, are described as CC terms. Example of CC terms include protein dimers, ribosomes, mitochondrion, and cell wall.

The distribution of enriched GO terms for upregulated and downregulated genes from *S. cerevisiae* cells encapsulated within each silica matrix is presented in **Figure 2.6**. In the Biological Process domain (**Figure 6, BP**), many GO terms are shared between all three methods for both up and down regulated genes. The CDA method showed more upregulated enriched GO terms (14) than

the other encapsulation methods (SD: 5; AqS+g: 1), with 9 shared GO terms between all three methods. For the downregulated genes, 25 GO terms are shared between three methods, however, the CDA method showed significantly more enriched BP terms (11) than the other two methods (SD: 4; AqS+g: 4). A significant majority of downregulated BP terms (14) are shared between the CDA and AqS+g methods. Also, cells encapsulated *via* CDA and SD shared 11 downregulated BP terms. In aggregate, these data show that all three methods shared more biological process terms than were unique to any one method of encapsulation. However, each of the encapsulation procedures produced some unique enriched classes.

In the Molecular Function GO domain (**Figure 2.6, MF**), most upregulated (8) and downregulated (12) terms are again shared between all three methods. Downregulated MF terms show little difference between the encapsulation methods with respect to number of unique MF terms (CDA:4; SD:8; AqS+g:1). These data indicate that AqS+g encapsulated *S. cerevisiae* cells show similar molecular function associated gene expression compared to cells in a stationary state.

In the Cellular Component GO domain (**Figure 2.6, CC**), the largest number of upregulated (11) and downregulated (11) CC GO enriched classes are shared by all three methods, followed by upregulated CC terms unique to CDA (8) and SD (7). Cells encapsulated in SD and CDA methods have a greater number of unique downregulated CC terms (6 and 5 respectively) than AqS+g, which has none.

Identification of Differentially Enriched GO Terms

Groupings of GO enrichment terms for Biological Process (BP) for encapsulated *S. cerevisiae* are presented in scatter plots by REVIGO[77] comprising **Figures 2.7, 2.9 and 2.10**. The scatter plots show the represented functional clusters after removing redundant terms by REVIGO.[77] The GO terms that are semantically similar should position together in the scatter plot and semantic space (X, Y), but it should be noted that X,Y coordinate units have no inherent meaning. The size of the circle is relative to the frequency of the GO term in the underlying GO database; the color of the circle represents the log 10 *p*-value. The threshold used for allowed sematic similarity is "medium." The corresponding complete listings of GO enrichment terms are provided in Appendix A, as are associated FDR-adjusted *p*-values for the GO enrichments.

In the following discussion, it is important to keep in mind that up- or down-regulation of expression of genes associated with a biological process may not be synonymous with up- or down-regulation of the process itself. Some of the genes may activate the process in question while others may inhibit it. Depending on overall feedback mechanisms, genes associated with activating a process may be upregulated to exploit availability of a substrate, while in other cases they may be upregulated to compensate for scarcity of a substrate. The only completely reliable inference to draw from enrichment of a gene ontology is with respect to association in some way with the process in question.

Enrichment of BP GO terms in each of encapsulation method is shown in **Figure 2.7**. All three methods show upregulated enrichment in cellular response to oxidative stress, proteolysis, oxidation-reduction process, protein transport and piecemeal microautophagy of nucleus (**Figure 2.7A**). Proteolysis is one of the key responses to stress as damaged proteins need to be eliminated from the cell.[24] Furthermore, reactive oxygen species (ROS) generated through oxidative phosphorylation can lead to chemical reactions that damage cells, which leads in turn to activation of oxidative stress responses from the cell.[24] All these processes are generally implicated in stress response.[24] Lipid and carbohydrate metabolic processes are upregulated in both SD and CDA (**Figure 2.7B, 2.7C, Appendix A respectively**). CDA and SD (but not AqS+g) share the upregulated BP term phosphorylation. Cells encapsulated *via* CDA showed upregulated cell wall organization terms (**Figure 2.7B**) as a response to weakened cell walls during stress. SD-encapsulated cells showed upregulated responses associated with obtaining energy by fatty acid beta oxidation (**Figure 2.7C**).

All three methods show downregulated enrichment in ribosome biogenesis, methylation, amino acid biosynthetic process, and a cluster of terms related to RNA processing (**Figure 2.7D-F, Appendix A**). Ribosome biogenesis and processing of RNAs such as tRNA or rRNA use a significant amount of energy and cells need to lower the expression of these genes to conserve energy under unfavorable conditions.[24] Reduction in expression level of ribosome biogenesis and RNA processing can be interpreted as a reduction in cell growth.[85] Furthermore, expression of genes associated with translation and tRNA aminoacylation are reduced as a general response to stress (**Appendix A**). Downregulation of these processes are characteristic features of stress

response.[24] Terms shared between CDA and SD, but not AqS+g, include metabolic process and phosphorylation. Both terms were also enriched in upregulated genes from CDA and SD, indicating cells from both methods have significantly changed their metabolic states and phosphorylation profiles in response to stress (**Figure 2.7B, 2.7C & 2.7E, 2.7F**).

BP GO terms that are unique to each of the three encapsulation methods (as mentioned in **Figure 2.6**) show that the different methods of encapsulation have some distinctly different effects on gene expression patterns (**Figure 2.8**). Upregulated genes in CDA are uniquely enriched in ascospore formation, fungal-type cell wall organization, aerobic respiration, glycogen biosynthetic process, and response to heat (**Figure 2.8A**). Ascospore formation in yeast is implicated in response to nutrition deprivation that allows cells to produce stress resistant haploid spores.[86] Furthermore, upregulated CDA genes are enriched in mitophagy and pexophagy (**Figure 2.8A**). Both mitophagy and pexophagy are selective autophagy processes to remove damaged mitochondria and peroxisomes, respectively in response to stress.[87,88] Upregulated genes in SD are uniquely enriched in protein targeting to mitochondrion, vesicle mediated transport, regulation of vacuole fusion and BP GO terms related to protein catabolism (**Figure 2.8A**). Upregulated genes in AqS+g are only enriched in the ubiquitin-dependent protein catabolic process *via* the multivesicular body sorting pathway (**Figure 2.8A**). Upregulation of ubiquitin dependent protein degradation is implicated in stress as damaged proteins are degraded by ubiquitination to adjust the cellular protein repertoire to cope with the new environment.[24]

Differentially downregulated genes in CDA are uniquely enriched in covalent chromatin remodeling, RNA splicing, protein transport, mRNA transport, mRNA processing, DNA replication, nucleoside metabolism and small metabolic processes, consistent with the observation that these processes are energy consuming and their expression levels are reduced under stress (**Figure 2.8B**).[24] Differentially downregulated genes in AqS+g are uniquely enriched in tRNA methylation, formation of translation preinitiation complex, transcription from RNA polymerase I promoter, and ribosomal subunit export from nucleus (**Figure 2.8B**), consistent with the observation of reduced levels of transcription and translation under stress.[24] Differentially downregulated genes in SD are uniquely enriched in amino acid transport, cell cycle, protein glycosylation and drug transmembrane transport (**Figure 2.8B**).

Enrichment of MF GO terms in each encapsulation method are arranged into clusters and shown in **Figure 2.9**. Upregulated genes in all three methods are enriched in oxidoreductase activity, lyase activity, peptidase activity, and structural constituents of cell wall terms (**Figure 2.9A-C**). Upregulation of oxidoreductase activity is consistent with upregulation of the BP term response to oxidative stress enriched in all three methods (**Figure 2.7A-C**). Downregulated genes from all three methods are enriched in translation initiation factor activity (**Figure 2.9D-F**). Both upregulated and downregulated genes from all three methods are enriched in nucleotide binding (**Figure 2.9A-F**). Downregulated clusters of aminoacyl-tRNA ligase, metal ion binding, and sequence-specific DNA binding are only enriched in CDA and SD (**Figure 2.9E, F**). Ligase activity and translation initiation factor binding are uniquely enriched in downregulated genes in the CDA method (**Figure 2.9E**). Transporter activity and antiporter activity are uniquely enriched in downregulated genes in the SD method (**Figure 2.9F**). A detailed list of unique differentially upregulated and downregulated MF terms in each method are displayed as heat map in **Supplementary Figure 2.1**.

Enrichment of CC GO terms in each encapsulation method is shown in **Figure 2.10**. The upregulated gene lists from all three methods are enriched in eisosome, cytoskeleton, and membrane terms (**Figure 2.10A-C**). Furthermore, upregulated and downregulated genes in all three methods are enriched in cytoplasm and polysome terms (**Figure 2.10A-F**). Upregulated genes in SD and AqS+g are both enriched in proteasome complex (**Figure 2.10C, Appendix A** respectively). The downregulated genes in both CDA and SD are downregulated enriched in nucleoplasm, and plasma membrane (**Figure 2.10E, 2.10F**). In contrast to the SD and AqS+g gene sets, the upregulated genes in CDA method are enriched in the Cellular Component categories fungal-type cell wall, extracellular region, and anchored component of membrane terms (**Figure 2.10B**). These are consistent with cell wall-associated observations in the BP enriched GO terms in same (CDA) method (**Figure 2.7B**). Furthermore, the CC term "cellular bud" is only enriched in cells encapsulated by the CDA method (**Figure 2.10B**). A detailed list of unique differentially upregulated and downregulated CC terms in each method are displayed as a heatmap in **Supplementary Figure 2.2**.

Enriched KEGG pathways (to a *p*-value of <0.01) are shown for upregulated and downregulated

genes for each encapsulation method in **Tables 2.1 and 2.2**, respectively. For all three encapsulation methods, metabolic pathways, biosynthesis of secondary metabolites, endocytosis, carbon metabolism, peroxisome, regulation of autophagy, and protein processing in endoplasmic reticulum are enriched in upregulated genes. Upregulated genes of AqS+g are uniquely enriched in fatty acid degradation, pyruvate metabolism, biosynthesis of amino acids, and tryptophan metabolism pathways. Upregulated genes of SD are uniquely enriched in proteasome and methane metabolism pathways. In the case of CDA, upregulated genes are uniquely enriched in meiosis, starch and sucrose metabolism, and phosphatidylinositol signaling system pathways.

In the case of downregulated genes, ribosome biogenesis in eukaryotes, metabolic pathways, pyrimidine metabolism, purine metabolism, RNA transport, Aminoacyl-tRNA biosynthesis, RNA polymerase, biosynthesis of secondary metabolites, and biosynthesis of amino acids are enriched in all three methods. Downregulated genes of CDA are uniquely enriched in RNA degradation, ribosome, spliceosome, arginine biosynthesis, and the mRNA surveillance pathway. Downregulated genes in SD are uniquely enriched in meiosis, glycine, serine and threonine metabolism, and the MAPK signaling pathway. A more extensive table showing all enriched pathways at p-value <0.05 is provided in **Appendix A**.

Unique Metabolic States for Encapsulated *S. cerevisiae* Cells

These enrichment results provide substantial new insight and understanding regarding the physiological state of living cells confined within silica matrices. Each of the three matrices studied in this work induced a unique metabolic state in the encapsulated *S. cerevisiae* cells. Significant overlap in gene expression profiles was observed between CDA and SD entrapment methods, which may be expected as both methods employ identical silica-lipid sol precursor solution chemistry. Still, substantial differences in gene expression profiles between the two methods were observed. In the case of CDA, cell wall processes were uniquely upregulated. This indicates that of the three encapsulation methods, cells entrapped *via* CDA experience greater cell wall associated stresses. This may be a result of the higher exposure time of these cells to the precursor sol solution, which contains ethanol, acid and hydrolyzed silica particles that may interact with, or pass through, the cell wall and stress or damage the cell. *S. cerevisiae* cells use the cell wall integrity (CWI) pathway to maintain cell wall integrity and repair damage *via* cell wall

biosynthesis and actin organization.[89] The cell wall integrity pathway also cross-talks to other cell stress pathways to relieve stresses.[90] Cells entrapped *via* CDA also showed substantial reduction in gene expression associated with ribosome biosynthesis and RNA processing. Biosynthetic processes require considerable energy, and reduced expression of rRNA processing, and ribosome biogenesis indicate that the cells are under stress and are minimizing energy consumption.

For cells encapsulated *via* the SD process, fatty acid beta oxidation GO term was very highly enriched (FDR-adjusted *p*-value=0.00001) in upregulated genes. The upregulation of these terms indicates that in the absence of a carbon nutrient source, *S. cerevisiae* cells can generate energy from the oxidation of fatty acids present in the encapsulation matrix at the cell surface. This results in acetyl moieties that can be used in the TCA cycle for generation of energy through anabolic metabolism. Interestingly, oxidation of fatty acids occurs in peroxisomes of *S. cerevisiae* cells,[91] and this is in agreement with the upregulated CC GO peroxisomal matrix term (**Appendix A**). As indicated previously, and as observed in previous reports, our results suggest that there is a lipid interface between the cell and the surrounding lipid-templated matrix, which may offer the cell a source of nutrients. We previously showed using a fluorescence recovery after photobleaching (FRAP) assay that this lipid layer is interdigitated throughout the entire bulk of CDA thin films and SD particles, and it remains extremely fluid. Thus, encapsulated cells are presented with a replenishing source of fatty acid nutrients even in the dry particle state. SD method entrapped cells also show significant reduction in ribosomal biogenesis and rRNA processing, indicating that the cells are under stress, but not to the same extent as cells entrapped *via* CDA. This may allow for some metabolic load to be focused on lipid metabolism for energy generation.

Overall, AqS+g cells exhibited fewer differentially expressed genes and fewer differentially expressed GO terms than either CDA or SD, indicating that this encapsulation method induced less drastic disruption than either SD or CDA. In particular, S. *cerevisiae* AqS+g cells lacked stress-related terms exhibited by CDA, or SD, or both including: upregulated lipid and carbohydrate metabolism such as observed in CDA and SD, upregulated cell wall organization and downregulation of mRNA transport, mRNA processing, nucleoside metabolism, and mitophagy observed in CDA. Furthermore, AqS+g cells use ubiquitin dependent protein catabolism, while SD cells use proteasomal machinery for protein catabolism (*i.e.* ubiquitin independent protein

catabolism) along with ubiquitin dependent protein catabolism. Overall, by the measure of stress-related enriched GO categories, the AqS+g method encapsulated cells are less stressed than cells entrapped *via* CDA or SD. We believe one reason is the inclusion of glycerol in AqS+g matrix, which is known to enhance long-term cellular viability.[20] This is attributed to the formation of a high glycerol and water containing region between the cells and the surrounding nanostructured silica matrix, resulting in a fluid interface between the cell membrane and polar silanol groups that may damage the membrane, while also insulating the cell from stresses induced during gelation.[19,33] We further believe that *S. cerevisiae* cells encapsulated in AqS+g can use glycerol as a source of carbon nutrient.[92]

Quiescent states in *S. cerevisiae* are commonly induced by stress, such as starvation, anoxia, *etc*.[84] These states are generally characterized by downregulation of genes associated with many forms of biosynthesis, but an upregulation of genes associated with forms of maintenance, for example cell wall maintenance.[93] Within this general framework, we note that *S. cerevisiae* does not exhibit one, but rather several different-quiescent states depending on the particular stress that induced the state (starvation of a particular nutrient, for example).[93] Gene expression patterns were further analyzed and compared to known, naturally occurring metabolic states. Major states identified in the literature include exponential growth, stationary, quiescent, dormancy, and persister.[6,84,94] It is difficult to distinguish between the states that various authors describe as quiescence, dormancy, and persistence, to the extent that it may be reasonable to consider them different names for the same entity. In the remainder of this paper we will adopt the term "quiescence," not because we prefer it to the other two, but simply to have one label.

The gene expression patterns induced by all three methods are generally consistent with the pattern of quiescence, but as with other stresses there are differences in the details of the quiescent states. The above mentioned major shifts of gene expression we see in encapsulated cells relative to the unencapsulated stationary state are characteristic of the quiescent state. In general, quiescent states are phenotypic variants of wild type cells and are induced by stresses such as nutrient starvation, antifungal stress or induction of biofilm formation.[95] Under conditions of high nutrient availability, the TOR complex 1 (TORC1) pathway plays an important role in cell growth by positively regulating ribosome biogenesis, and inhibiting stress associated pathways that have negative

effects on cell growth such as cell wall integrity (CWI) pathways.[89] However, under a stress environment or as a response to the unavailability of nutrients, TORC1 is inhibited, resulting in no cell growth, one characteristic of a quiescent state.[96] Bojsen *et al*[95] recently reported that inhibition of TORC1 in *S. cerevisiae* resulted in increased tolerance to antifungal agents.

Significance of Confinement and the Bio-Nano Interface on Living Cell Metabolic State

While confinement within silica matrices has been widely reported to induce significant and unexpected shifts in biological behaviors, a comprehensive understanding of the influence encapsulation matrix properties has on specific biological functions is lacking. In this first analysis of how systematic modifications of the encapsulating nanostructure influence *S. cerevisiae* cells genetic expression profiles, the physicochemical properties at the bio-nano interface influence the metabolic state of encapsulated cells beyond those induced by 3D physical confinement alone. This is evidenced by uniquely enriched genes and associated unique functional categories from each method. As shown in the heat map and Venn diagram in **Figure 2.5**, 750 upregulated, and 598 downregulated genes are shared between all three encapsulation chemistries. However, and importantly, there are a similar number of unshared up and down regulated genes among each encapsulation method (AqS+g: 444 up, 344 down; CDA: 565 up, 786 down; SD: 707 up, 694 down) This is also shown in **Figure 2.8**, and in **Supplementary Figures 2.1 and Supplementary Figure 2.2**. Physicochemical properties at the bio-nano interface are also more influential than the overall chemistry of the confinement matrix. Although both CDA and SD methods employ identical matrix chemistries and share many enriched genes (936 upregulated, 902 downregulated), many genes were still uniquely enriched by the two methods (CDA: 379 up, 482 down; SD: 521 up, 390 down). Overall, of the 3872 differentially regulated genes measured in this study, 1508, or 39%, were unique to a given encapsulation matrix, demonstrating that the interaction of the cell with the three differing nanostructures substantially impacts cell metabolism.

Here it is noteworthy that the three different encapsulating matrices were all amorphous silicate-based. There is a large body of work studying the toxicity of nanostructured materials, which has been shown to differ from the toxicity of bulk materials of the same chemical composition. That work shows that the nanostructure and type of nanomaterial clearly influence cellular metabolic

activity. From data collected in this study, any silica specific effects on the metabolic state of the cells is challenging to determine for. Still, it is known that amorphous silica materials are Generally Recognized as Safe (GRAS) by the US Food and Drug Administration, and that sol-gel derived silica nanoparticles have been FDA approved for diagnostic applications in a stage I human clinical trial.[97] It is therefore unlikely that there would be significant silica specific toxicity or effects on cell metabolism. In the case of CDA *versus* SD, the starting lipid/silica compositions were identical, the physicochemical properties at the bio-nano interface differed significantly. This includes differences in the physical interaction of lipid, glycerol, and silica with the cell wall, as shown in the TEM imaging data (**Figures 2.2-2.4**). Also, physical forces exerted at the bio-nano interface differ between each matrix as indicated by mechanical modulus data. As we recently reported, the Young's modulus of SD matrices is ~14 GPa, as determined by nano-indentation, compared to ~4 GPa for CDA determined in an identical manner.[17] Due to their porosity (see **Figure 2.4**), aqueous silica gel matrices are much softer. Based on literature reports, we estimate the Young's modulus of the AqS+g matrix to be ~ 500 KPa to 2MPa[80] (where Young's modulus, *E,* is calculated from the reported shear modulus, *G,* by $E = 2(1+n)\,G$, where Poisson's ratio n is assumed to be 0.2), over 1000x softer than SD. Thus, these differing matrices mechanically constrain the cell to differing extents, which could be manifested in multiple stress-associated pathways. Although the aqueous gels are quite soft, they still inhibited replication of *S. cerevisiae* as did the stiffer, conformal CDA and SD matrices. This contrasts with very soft silica gel matrices or hierarchical macro-meso-microporous silica gels where replication of *E. coli* was observed respectively.[64,69]

Therefore, when using materials chemistry to elicit a desired biological response, the bio-nano interface structural and chemical properties are highly significant parameters to control. This understanding opens the possibility of integrating living cells within functional materials that are less-biocompatible, or even cytotoxic, but may provide long-term viability and activity by incorporating a well-designed bio-nano interface. Finally, tuning the parameters of the bio-nano interface can induce yet unexplored cellular states that will require further study to understand, but may contribute substantial understanding of cancer, aging, cell signaling, quorum sensing, starvation, and other complex cellular behaviors.

Conclusions

In summary, we have performed a comprehensive genetic analysis of living cells encapsulated within silica matrices. Comparison of encapsulated cells RNA levels with stationary phase cells permitted the identification of stresses exerted on the cells associated with a given encapsulation chemistry, and a broad data set illustrative of the cells' metabolic processes. This will allow for further tuning of the given encapsulation approach to improve upon the desired biological traits, enhancing the performance of the hybrid biomaterial. This study also showed that the material properties at the bio-nano interface significantly influence biological behavior, in addition to the bulk chemistry of the matrix and physical confinement alone. Distinct metabolic states for cells entrapped in each silica matrix were observed. In future work, we believe it will be possible to induce, and perhaps control, a biological state by tuning the nanomaterial properties at the bio-nano interface of the encapsulation matrix. This ability may provide a powerful new technique for the study of complex cellular behaviors impacting the fields of cancer, aging, cell-cell signaling and quorum sensing research, and the development of bioelectronics and cell-based biosensors.

Figures

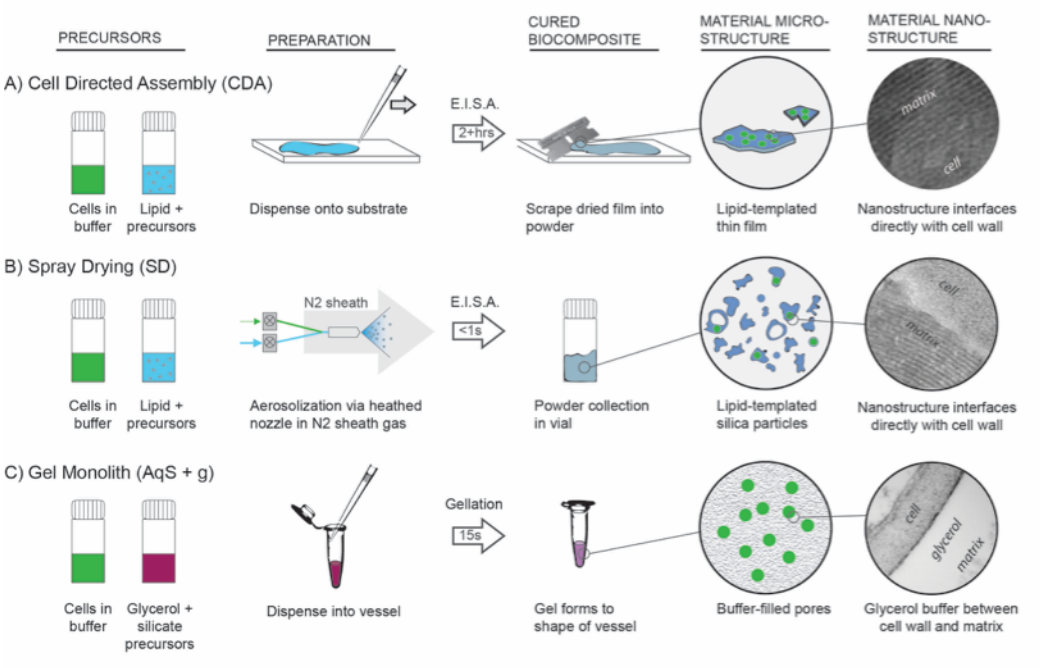


Figure 2.1. Schematic depicting the three cellular encapsulation methods investigated in this study. For all three methods, cells in buffer are mixed with precursor solution and either added to a substrate (A), spray dried (B), or added to a gelation vessel (C). In the first two cases, evaporation induced self-assembly (EISA) drives the formation of a nanostructured lipid-silica encapsulating matrix, which results in a dry film (A) or powder (B), both of which exhibit a conformal ordered nanostructure as visualized by TEM. In the third case, gelation occurs within the sealed vessel, physically entrapping the cells within a highly porous matrix composed of aggregated silica nanoparticles that does not conform to the cell surface.

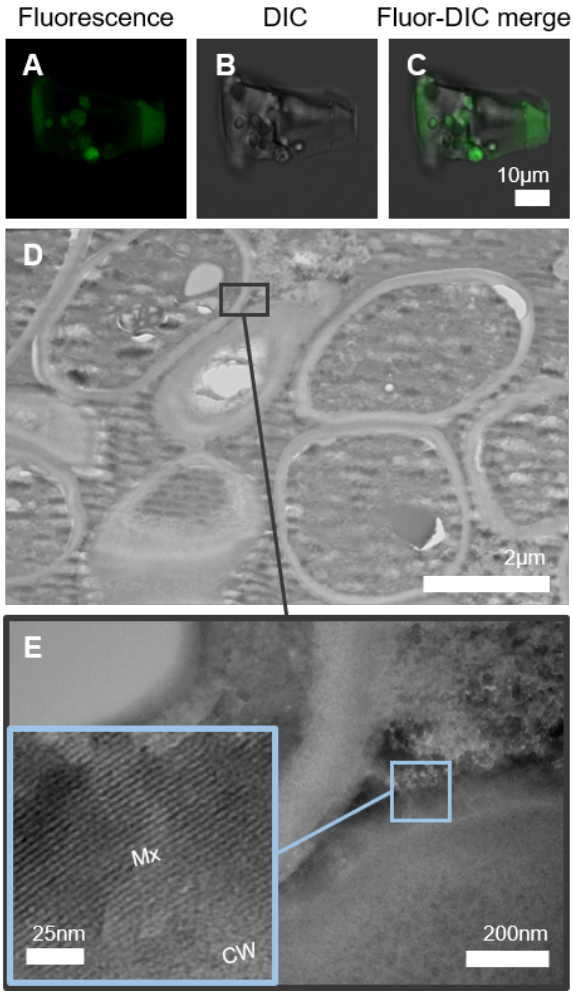


Figure 2.2. Micro- and nano-morphologies of CDA fragments show tightly packed cells. Fluorescence projection z-stack analysis verifies encapsulation while transmission electron microscopy (TEM) reveals an ordered nanostructured lipid/silica encapsulating matrix that interfaces with the cell surface. (A-C) Projection confocal fluorescence images of a small fragment highlight the placement of cells within the particle. (The green bands at the edge of the particle are an optical artifact resulting from refraction of the green fluorescence occurring within the glass-like particle). (D) TEM analysis shows a very dense cell loading within the particle and a zoomed image (E) reveals an ordered nanostructured matrix that, here, is highlighted between two cells. The inset clearly shows an ordered nanostructure interfacing between two cell walls. CW: cell wall; Mx: matrix.

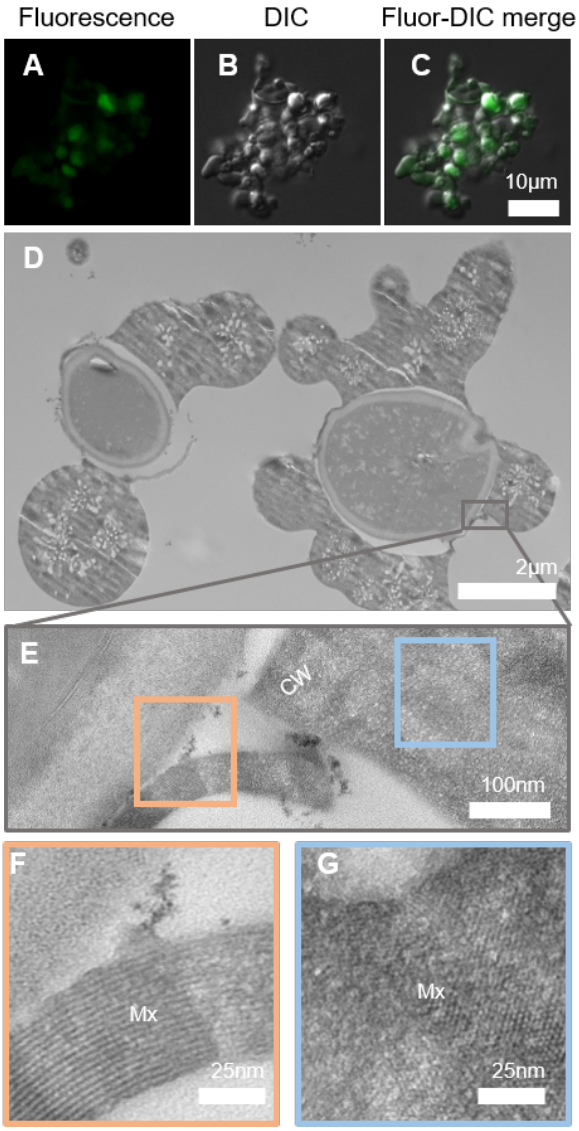


Figure 2.3. Micro- and nano-morphologies of SD particles reveal distribution of cells within dry, spray-dried powders. (A-C) Projection confocal fluorescence image (A), DIC image (B), and merged image (C) of a particle fragment highlight the placement of cells within the particle. (D) TEM analysis shows cells encapsulated within a conformal, ordered lipid-silica matrix in addition to whole lipid-silica particles, which contain no cells as described previously.[17] The chattering in the ultra-microtomed section is attributed to the high hardness and Young's modulus of the silica nanostructure as described previously.[17] (E) A zoomed image shows the interface of the lipid/silica nanostructure and the cell which appears to be "peeling off" from the cell surface. The nanostructure is clearly visible and interfaces directly with the cell surface (F) and extends into the bulk of the particle (G). CW: cell wall; Mx: matrix. Both the peeling and separation of the matrix from the cell are attributed to sample preparation artifacts given correspondence of the shape of the separated matrix and cell surface.

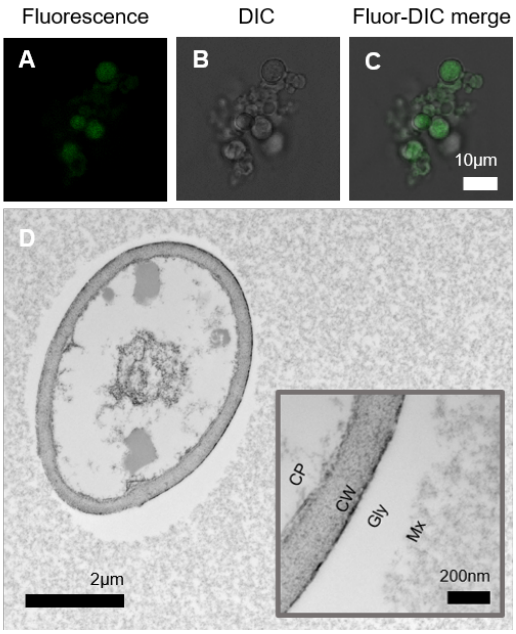


Figure 2.4. Micro- and nano-morphologies of gel particles show cells that are individually incorporated within an amorphous matrix and are surrounded by a glycerol sheath. (A-C) Projection confocal fluorescence images of a small particle highlight the placement of cells within the particle. (D) TEM images of an individual cell show incorporation within a ramified, highly porous matrix composed of small (10 nm) primary silica nanoparticles that is separated from the cell surface by a 20-450 nm thick porous layer that is rich in glycerol. CP: cytoplasm; CW: cell wall; Gly: glycerol; Mx: matrix.

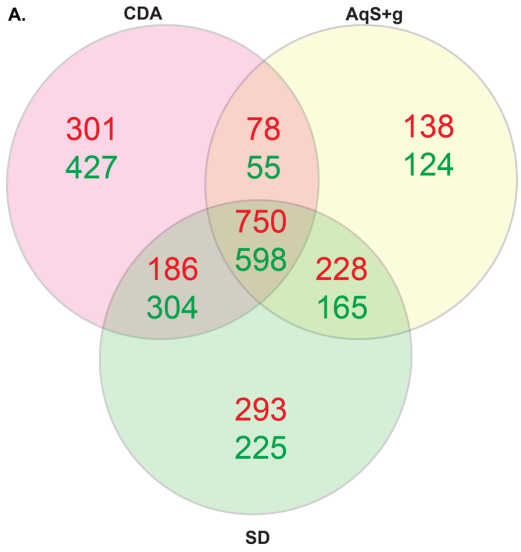


Figure 2.5. Distribution of all differentially expressed genes (FDR-adjusted p-value <0.05, (|log2(fold change) | > 1) from *S. cerevisiae* cells encapsulated within three differing silica matrices: Cell-directed assembly (CDA); spray drying (SD); aqueous silicate with glycerol (AqS+g). Upregulated genes are shown in red. Downregulated genes are shown in green.

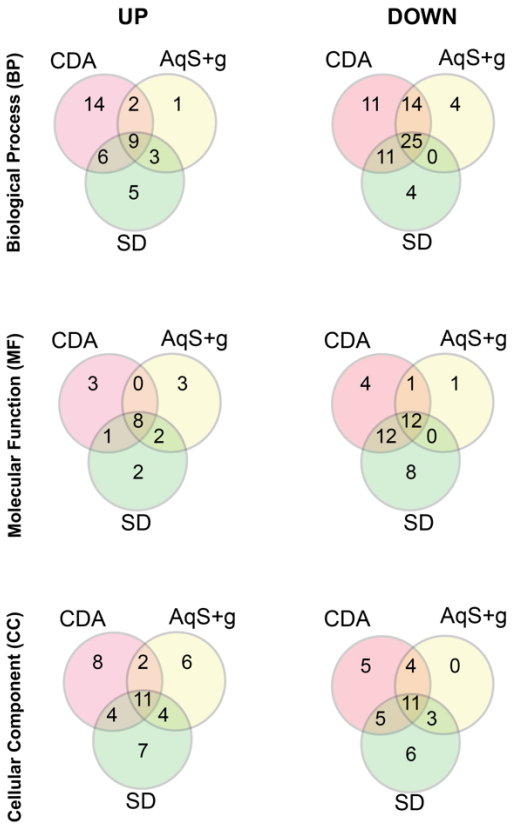


Figure 2.6. Gene ontology (GO) categorization of differentially expressed genes from silica matrix encapsulated S. cerevisiae cells. GO domains: Biological Process (BP); Molecular Function (MF); Cellular Component (CC). Silica encapsulation matrices: Cell-directed assembly (CDA); Spray drying (SD); Aqueous silicate with glycerol (AqS+g).

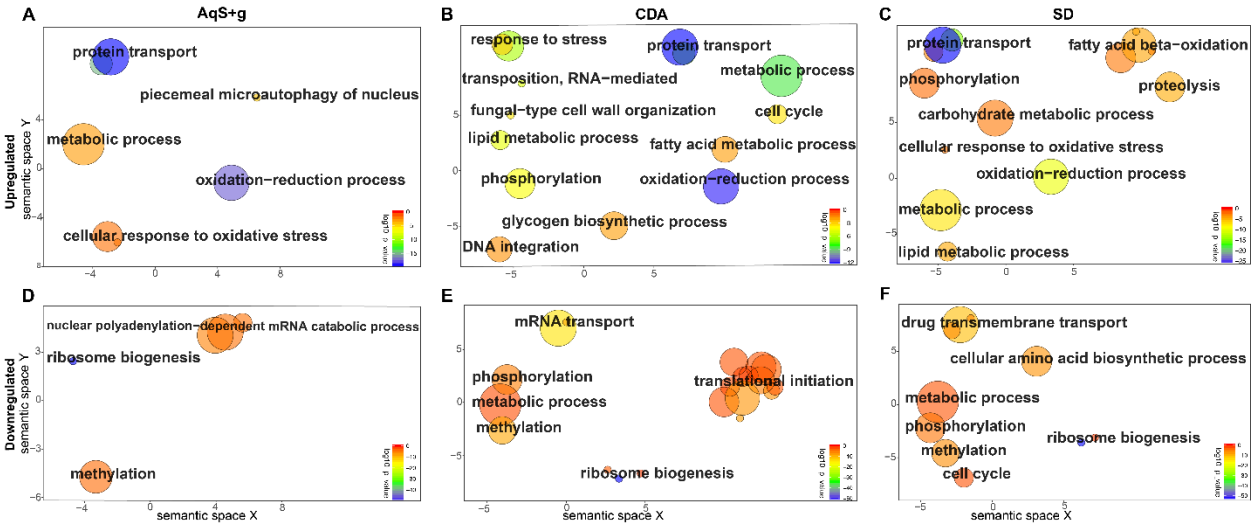


Figure 2.7. Gene Ontology (GO) enrichment of biological processes (BP) terms as determined according to The Database for Annotation, Visualization and Integrated Discovery (DAVID). Redundant GO terms were removed and then summarized by REVIGO and represented as scatter plot. (A-C) Upregulated biological processes for *S. cerevisiae* encapsulated *via* (A) AqS+g, (B) CDA, (C) SD. (D-F) Downregulated biological processes for *S. cerevisiae* encapsulated *via* (D) AqS+g, (E) CDA, (F) SD. GO terms are shown by circles and plotted by semantic similarity with other GO terms, circles closer to each other shows closely related GO terms. The size of the circle is relative to the frequency of the GO term in the underlying GO database. Circle color represents the log 10 *p*-value. GO terms with FDR-adjusted *p*-value <0.05 are shown.

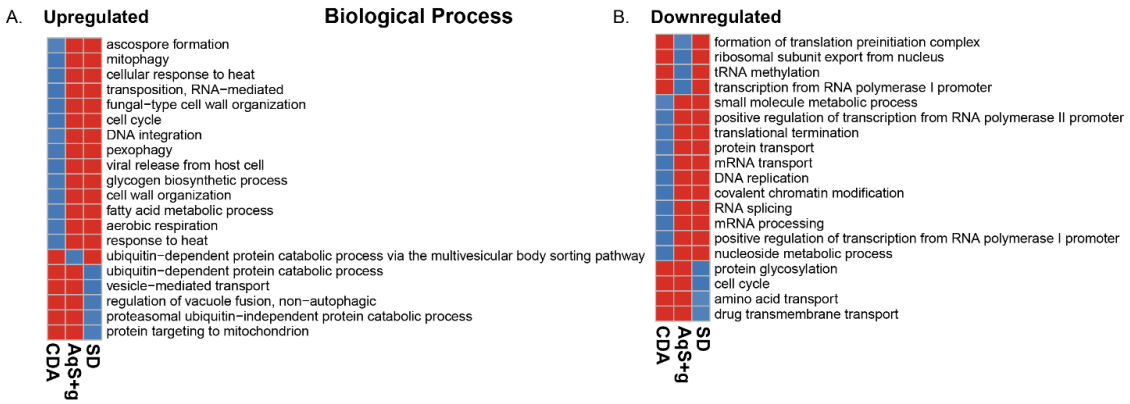


Figure 2.8. Biological Process (BP) Gene ontology (GO) terms that are uniquely enriched in each of the three encapsulation methods. GO enrichment was determined *via* the Database for Annotation, Visualization and Integrated Discovery (DAVID). (A) Upregulated and (B) downregulated BP for *S. cerevisiae* encapsulated *via* AqS+g, CDA, and SD. Red colored boxes show absence of a GO term, while blue color box show presence of a GO term in the given encapsulation method. GO terms with FDR-adjusted p-value < 0.05 are shown. Individual FDR-adjusted p-values are shown in Appendix A.

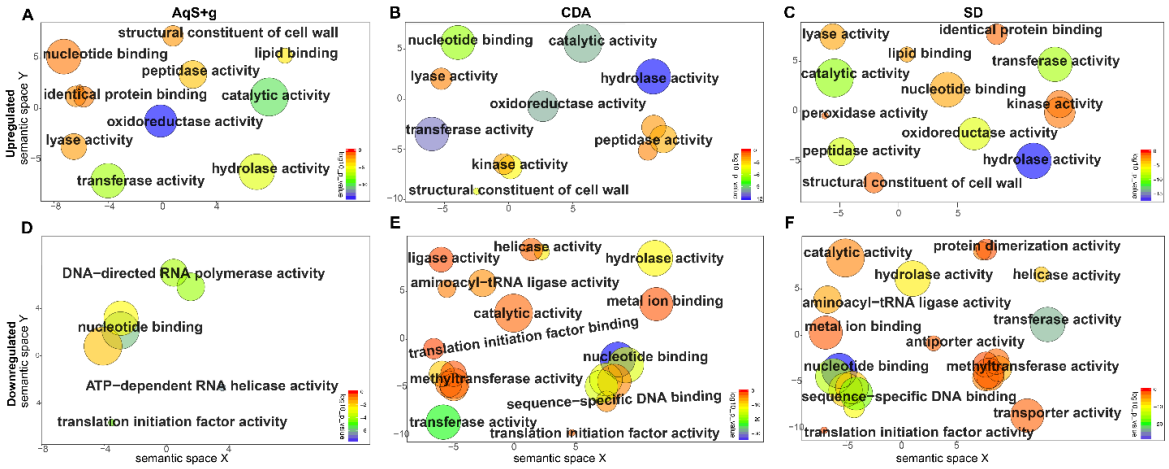


Figure 2.9. Gene Ontology (GO) enrichment of molecular function (MF) terms as determined according to the Database for Annotation, Visualization and Integrated Discovery (DAVID). Redundant GO terms were removed and then summarized by REVIGO and represented as scatter plot. (A-C) Upregulated molecular functions for *S. cerevisiae* encapsulated *via* (A) AqS+g, (B) CDA, (C) SD. (D-F) Downregulated molecular functions for *S. cerevisiae* encapsulated *via* (D) AqS+g, (E) CDA, (F) SD. GO terms are shown by circles and plotted by semantic similarity with other GO terms, circles closer to each other show closely related GO terms. The size of the circle is relative to the frequency of the GO term in underlying GO database. Circle color represents the log 10 *p*-value. GO terms with FDR-adjusted *p*-value <0.05 are shown.

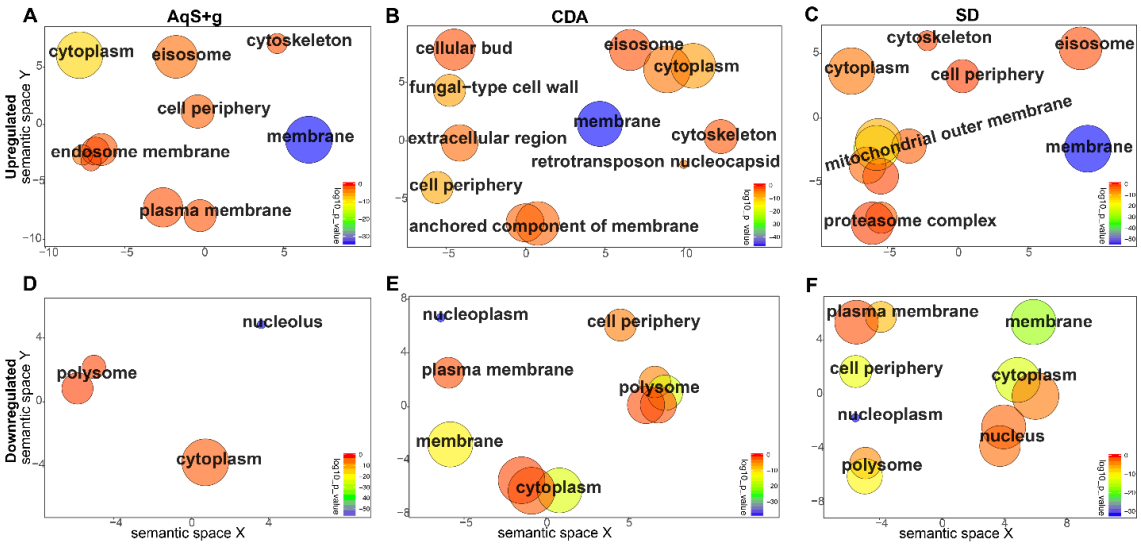


Figure 2.10. Gene Ontology (GO) enrichment of Cellular Component (CC) terms as determined according to The Database for Annotation, Visualization and Integrated Discovery (DAVID). Redundant GO terms were removed and then summarized by REVIGO and represented as scatter plot. (A-C) Upregulated cellular components for *S. cerevisiae* encapsulated *via* (A) AqS+g, (B) CDA, (C) SD. (D-F) Downregulated cellular components for *S. cerevisiae* encapsulated *via* (D) AqS+g, (E) CDA, (F) SD. GO terms are shown by circles and plotted by semantic similarity with other GO terms, circles closer to each other shows closely related GO terms. The size of the circle is relative to the frequency of the GO term in underlying GO database. Circle color represents the log 10 *p*-value. GO terms with FDR-adjusted *p*-value <0.05 are shown.

Tables

Table 2.1 Enriched Pathways in Upregulated Genes in Each Encapsulation Method.

Pathway Name	FDR-adjusted *p*-value		
	CDA	**AqS+g**	**SD**
Metabolic pathways	5.33E-20	5.33E-17	3.32E-23
Biosynthesis of secondary metabolites	5.04E-10	1.25E-12	1.89E-13
Endocytosis	4.63E-05	8.79E-12	5.97E-08
Carbon metabolism	1.17E-06	6.51E-07	2.30E-07
Peroxisome	5.32E-05	8.58E-07	4.74E-07
Protein processing in endoplasmic reticulum	4.22E-05	1.78E-04	5.60E-05
Glycerolipid metabolism	>0.01	1.99E-04	1.94E-04
Fatty acid degradation	>0.01	5.31E-04	>0.01
Pyruvate metabolism	>0.01	9.05E-04	>0.01
SNARE interactions in vesicular transport	>0.01	0.0011	9.82E-04
Regulation of autophagy	3.54E-05	0.0011	1.07E-04
Glycolysis / Gluconeogenesis	0.0016	0.0025	>0.01
Biosynthesis of amino acids	>0.01	0.0027	>0.01
Tryptophan metabolism	>0.01	0.0065	>0.01
Proteasome	>0.01	>0.01	2.46E-05
Methane metabolism	>0.01	>0.01	0.0068
Glycerophospholipid metabolism	0.0032	>0.01	0.0035
Meiosis	3.38E-06	>0.01	>0.01
Starch and sucrose metabolism	9.77E-06	>0.01	>0.01
Phosphatidylinositol signaling system	0.0066	>0.01	>0.01

Table 2.2 Enriched Pathways in Downregulated Genes in Each Encapsulation Method.

Pathway Name	FDR-adjusted *p*-value		
	CDA	**AqS+g**	**SD**
Ribosome biogenesis in eukaryotes	3.25E-26	6.96E-24	1.50E-25
Metabolic pathways	4.77E-22	8.16E-13	2.60E-20
Pyrimidine metabolism	3.11E-12	1.27E-09	2.21E-11
Purine metabolism	3.49E-12	6.84E-08	1.00E-08
RNA transport	1.85E-14	2.66E-06	2.13E-08
Aminoacyl-tRNA biosynthesis	1.25E-04	1.10E-05	1.00E-04
RNA polymerase	2.04E-09	5.11E-04	2.35E-08
Biosynthesis of secondary metabolites	1.40E-07	5.55E-04	1.41E-07
Biosynthesis of amino acids	4.65E-08	5.72E-04	2.26E-10
Phenylalanine, tyrosine and tryptophan biosynthesis	>0.01	0.0027	0.0022
Cell cycle - yeast	0.0034	>0.01	3.45E-04
RNA degradation	4.23E-05	>0.01	>0.01
Ribosome	2.70E-04	>0.01	>0.01
2-Oxocarboxylic acid metabolism	0.0017	>0.01	0.0015
Cysteine and methionine metabolism	0.0023	>0.01	5.70E-04
Spliceosome	0.0029	>0.01	>0.01
Alanine, aspartate and glutamate metabolism	0.0052	>0.01	0.0070
Arginine biosynthesis	0.0084	>0.01	>0.01
mRNA surveillance pathway	0.0089	>0.01	>0.01
Meiosis - yeast	>0.01	>0.01	7.27E-06
Glycine, serine and threonine metabolism	>0.01	>0.01	5.63E-04
MAPK signaling pathway - yeast	>0.01	>0.01	0.0025

Table 2.3 Gene Set Enrichment Analysis Categories over-represented and under-represented in AqS+g relative to control (Nominal *p*-value =0)

GO Term	NG	p-value	FDR
Over-represented			
protein targeting to vacuole & establishment of protein localization to vacuole	91	0	0.144
protein localization to vacuole	92	0	0.170
establishment of protein localization to organelle	276	0	0.244
protein localization to organelle	332	0	0.249
intracellular protein transport	319	0	0.308
protein transport	344	0	0.316
mitochondrion organization	281	0	0.319
protein targeting	271	0	0.323
vacuolar transport	142	0	0.371
establishment of protein localization	370	0	0.446
Under-represented			
preribosome	170	0	0.003
septin ring organization	28	0	0.018
septin cytoskeleton organization	29	0	0.018
maturation of SSU rRNA from tricistronic rRNA transcript SSU rRNA 5.8s rRNA LSU rRNA	96	0	0.022
maturation of SSU rRNA	105	0	0.023
rRNA metabolic process	241	0	0.023
nucleolus	245	0	0.024
ribonucleoside triphosphate metabolic process	43	0	0.024
purine containing compound biosynthetic process	51	0	0.026
90s preribosome	85	0	0.029

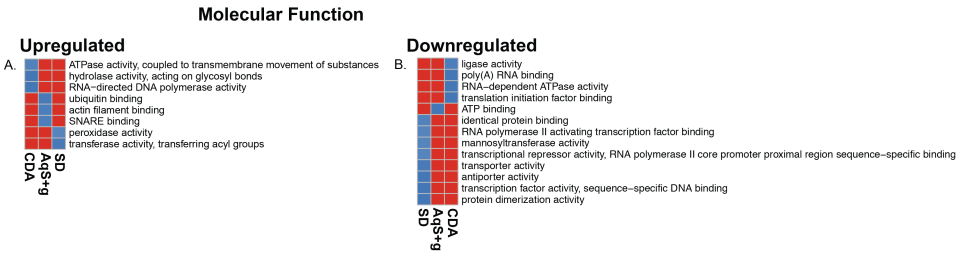
Table 2.4 Gene Set Enrichment Analysis Categories over-represented and under-represented in CDA relative to control (Nominal *p*-value =0)

GO Term	NG	p-value	FDR
Over-represented			
external encapsulating structure & cell wall & fungal type cell wall	89	0	0.023
hydrolase activity hydrolyzing on glycosyl compounds	35	0	0.024
mitochondrion localization	35	0	0.064
cortical endoplasmic reticulum	17	0	0.112
energy reserve metabolic process	34	0	0.112
protein localization to vacuole	92	0	0.112
CVT pathway	37	0	0.116
monovalent inorganic cation homeostasis	40	0	0.118
hydrolase activity acting on glycosyl bonds	43	0	0.123
extracellular region	27	0	0.160
Under-represented			
ncRNA metabolic process	435	0	0
ribosome biogenesis	327	0	0
preribosome	170	0	0
ncRNA processing	327	0	0
nucleolus	245	0	0
ribonucleoprotein complex biogenesis	402	0	0
rRNA metabolic process	241	0	0
ribosomal large subunit biogenesis	93	0	0
ribosomal small subunit biogenesis	125	0	0
rRNA processing	224	0	0

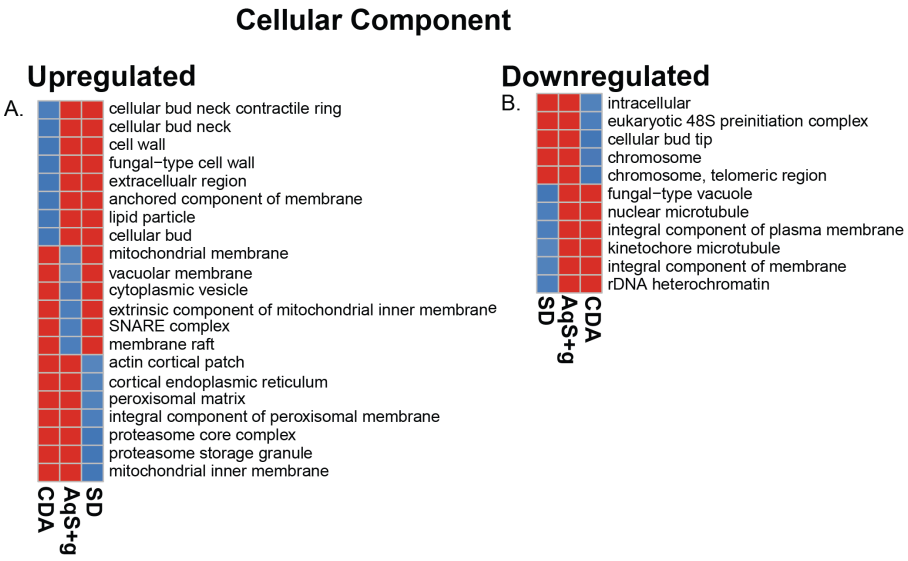
Table 2.5 Gene Set Enrichment Analysis Categories over-represented and under-represented in SD relative to control (Nominal *p*-value =0)

GO Term	NG	p-value	FDR
Over-represented			
pre autophagosomal structure	22	0	0.156
mitochondrial outer membrane	93	0	0.182
protein targeting to vacuole & establishment of protein localization to vacuole	91	0	0.186
positive regulation of protein complex assembly	39	0	0.190
extrinsic component of membrane	92	0	0.204
microbody & peroxisome	69	0	0.205
mitochondrion organization	281	0	0.215
glycogen metabolic process	32	0	0.221
outer membrane & organelle outer membrane	96	0	0.266
vacuolar transport	142	0	0.297
Under-represented			
nucleolus	245	0	0
ribosome biogenesis	327	0	0
preribosome	170	0	0
rRNA processing	224	0	0
ncRNA processing	327	0	0
ribonucleoprotein complex biogenesis	402	0	0
ncRNA metabolic process	435	0	0
rRNA metabolic process	241	0	0
ribosomal small subunit biogenesis	125	0	0
maturation of SSU rRNA	105	0	0

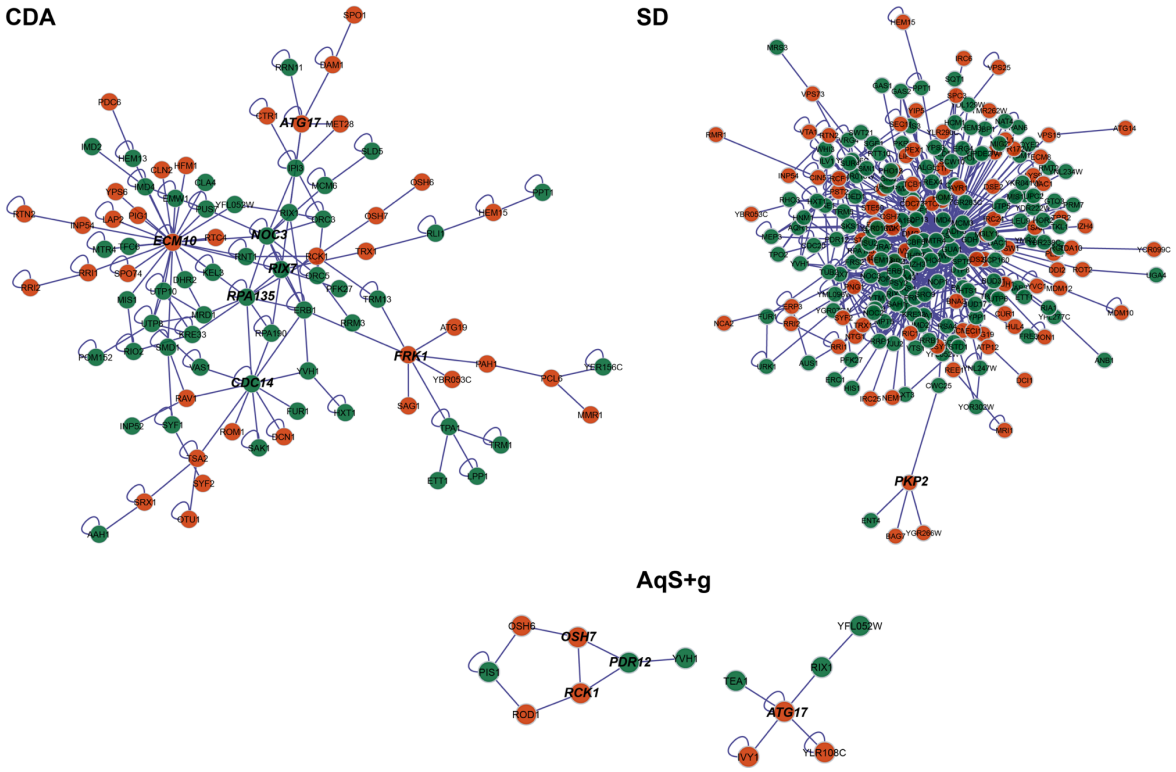
Supplementary Figures



Supplementary Figure 2.1. Molecular Function (MF) Gene ontology (GO) terms that are uniquely enriched in each of the three encapsulation methods. Gene Ontology (GO) enrichment was done *via* The Database for Annotation, Visualization and Integrated Discovery (DAVID). (A) Upregulated molecular functions for S. cerevisiae encapsulated via AqS+g, CDA, and SD. (B) Downregulated molecular functions for S. cerevisiae encapsulated via AqS+g, CDA, and SD. Red color box show absence of a GO term while blue color box show presence of a GO term in encapsulation method. GO terms with FDR-adjusted *p*-value of 0.05 or less are shown. Individual FDR-adjusted *p*-values are shown in Appendix A.



Supplementary Figure 2.2. Cellular Component (CC) Gene ontology (GO) terms that are uniquely enriched in each of the three encapsulation methods. Gene Ontology (GO) enrichment was done *via* The Database for Annotation, Visualization and Integrated Discovery (DAVID). (A) Upregulated cellular component for S. cerevisiae encapsulated via AqS+g, CDA, and SD. (B) Downregulated cellular component for S. cerevisiae encapsulated via AqS+g, CDA, and SD. Red color box show absence of a GO term while blue color box show presence of a GO term in encapsulation method. GO terms with FDR-adjusted *p*-value of 0.05 or less are shown. Individual FDR-adjusted *p*-values are shown in Appendix A.



Supplementary Figure 2.3. Network of differentially expressed genes (FDR-adjusted p-value <0.05, and (|log2 fold change|) >5 in each encapsulation method (CDA, SD, and AqS+g). Green Node color indicated downregulated genes, Red color indicated upregulated genes. Edges show the relationship between gene from databases in BisoGenet. Interaction between differentially expressed genes is shown.

Chapter 3: Time-Varying Response of Yeast Cells to Encapsulation in a Silica Matrix and Subsequent Reincubation in a Growth Medium

Abstract

We report on the temporal pattern of yeast gene expression patterns during encapsulation in silica matrices via a cell-directed assembly process, and upon release. Three broad classes of patterns are seen. A major shift in expression patterns is seen upon encapsulation, relative to the beginning stationary state, similar to previously reported stress response. Significant continuing shifts are seen by sampling at different intervals during a one week encapsulation. Upon release from encapsulation and reincubation in growth medium, the cells are in a state significantly different from the state prior to encapsulation and similar to the state during encapsulation. Implications are drawn for the use of encapsulated micro-organism as sensors and effectors, and for the persister state of such organisms.

Introduction

Encapsulation of microbial cells can disturb the cellular environment and impose various stresses. Encapsulation stresses include mechanical stress, tensile stress, acid exposure, starvation stress, and temperature shocks.[57–59] To handle these environmental changes, cell must adjust internal system to perform regular functions. One of the yeast *Saccharomyces Cerevisiae (S. Cerevisiae)* responses to these changes is the provisional pause of normal cellular processes during the reorganization of the internal system.[3] In response to stress, cells often change their physiological or metabolic state to induce the expression of those genes who are involved in cellular processes required for growth and survival. Microbial cells present in biofilms can change their metabolic state to enhance antibiotic resistance/tolerance and formation of persister cells.[98,99] Furthermore, encapsulation in nanostructure silica matrix has induced quorum sensing in individual *Staphylococcus* aureus cells.[66] The response to many stresses like deprivation of nutrients, growth at high or low temperature, osmolarity, acidic environment and presence of toxic agents have been studied.[3] In a prior study we explored the stationary response of yeast cells to various methods of encapsulation in silica matrices.[100] In the present study, we present and analyze the time-varying gene expression patterns of *S. cerevisiae* cells encapsulated in a cell-directed assembly (CDA)

method of encapsulation, and compare our results to published results of yeast cells under other types of stress.

Material and Methods

Materials

Saccharomyces cerevisiae (strain S288c) was acquired from American Type Culture Collection (ATCC; Manassas, VA). 1,2-dihexanoyl-sn-glycero-3-phosphocholine (*di*C_6 PC) was purchased from Avanti Polar Lipids, Inc. (Alabaster, AL). Absolute ethanol, hydrochloric acid (HCl, 37%), tetraethylorthosilicate (TEOS, 98%), sodium silicate solution (26.5% SiO_2; 10.6% Na_2O), strongly acid cation-exchange resin (DOWEX 50WX8-100, hydrogen, 50-100 mesh), glycerol, yeast extract, peptone, D-(+)-glucose, adenine, sodium acetate, ethylenediaminetetraacetic acid (EDTA), sodium dodecyl sulfate (SDS), phenol, choloroform:isoamyl alcohol (24:1), sodium phosphate (mono- and dibasic), sodium acetate, phosphate buffer saline (PBS, pH 7.0), glass beads, and Alconox detergent were purchased from Sigma-Aldrich (St. Louis, MO). RNeasy Mini Kit (including Buffer RW1 and Buffer RPE) was obtained from Qiagen (Germantown, MD). RNA 6000 Nano Kit was purchased from Agilent (Santa Clara, CA). Yeast Genome 2.0 Arrays and Genechip 3' IVT Express Kits were from Affymetrix (Santa Clara, CA).

Cell Culture

S. cerevisiae S288c cells were grown aerobically in Yeast Peptone Dextrose + Adenine (YPD+A) broth at 30 °C for seven days until the culture was in stationary phase (OD600 1.5-2.0). Cells were then centrifuged at 10,000 rpm for 5 mins, washed with DI H_2O three times, and re-suspended in DI H_2O (OD600 0.1-1.0).

Encapsulation of S. cerevisiae in Phospholipid-Templated Silica Films via Cell Directed Assembly (CDA)[18]

Glass coverslips (25-mm, No. 1.5) were soaked in 0.1 M KOH for a minimum of 4hr, washed with 10% Alconox (m/v), rinsed with DI H_2O, dried under a stream of N_2 gas, and then cleaned in UV-ozone (Jelight, model 342) for 5mins. Prehydrolyzed tetraethyl orthosilicate (TEOS) stock solutions (A2**) were made by refluxing 61 mL of TEOS, 61 mL of absolute ethanol, 4.9 mL of DI H_2O, and 0.2 mL of 0.07 N HCl (molar ratio $1:4:1:5x10^{-5}$) for 90mins at 60°C. A2** stock

solutions were then stored at -20°C. CDA precursor sol was prepared by combining 0.25 mL of A2** stock solution, 0.2 mL of absolute ethanol, 0.16 mL of 0.05 N HCl, 0.4 mL of DI H_2O and 30 mg of *di*C_6 PC. This precursor solution was aged at room temperature for 20mins and then passed through a 0.45-μm filter. To this solution was combined an equal volume of PBS washed and re-suspended *S. cerevisiae*. This mixture was cast onto cleaned glass coverslips. Resulting thin films could air dry, were rinsed with DI H_2O to remove non-integrated cells, and stored under ambient conditions (25˚C ± 2˚C) for three days.

Extraction and Purification of RNA from Encapsulated Cells

Portions of encapsulated cell-containing silica matrices were rinsed with DI H_2O and placed in 2 mL centrifuge tubes, breaking the matrix into smaller pieces as necessary. 270 μL sodium acetate buffer (50 mM sodium acetate, pH 5.2, and 10 mM EDTA, pH 8.0), 30 μL 10% SDS, and approximately 100 μL of glass beads (0.5 – 0.6 mm diameter) were added to centrifuge tubes. Samples were stirred and agitated to physically break apart the matrix. Water saturated acid phenol (300 μL) preheated to 65°C was added to the samples, followed by vortexing for 1min. Samples were then incubated in a 65°C water bath for 5mins, followed again by vortexing for 1min. This was repeated for a total of six cycles over approximately 45mins. Samples were then placed in a cooling block for 2mins, followed by the addition of 300 μL chloroform:isoamyl alcohol (24:1), and mixed by vortexing for 30 seconds. Samples were then centrifuged for 5mins at 15,000 rpm, followed by transferring the RNA-containing top layer to new centrifuge tubes. An equal volume of 70% ethanol was added and mixed by pipetting.

Next, samples were transferred to Qiagen RNeasy spin columns and centrifuged for 30 seconds at 15,000 RPM, discarding flow through. 700 μl of RW1 buffer was added to the columns and the samples were centrifuged for 30 seconds at 15,000 RPM, discarding flow through. 500 μL of RPE buffer was added to the columns, followed by centrifugation for 30 seconds at 15,000 RPM, discarding flow through. 500 μL of RPE buffer was added to the columns and followed by centrifuging for 2mins at 15,000 RPM, discarding flow through. Columns were then placed in new 2 mL collection tubes and centrifuged for 1min at 15,000 RPM. Next, columns were placed in new 1.5 mL collection tubes and 40 μL of DI H_2O was added. Columns were again centrifuged for 1min at 15,000 RPM, after which 40 μL of DI H_2O was added and the columns were centrifuged

again for 1min, yielding the purified RNA product. A NanoDrop spectrophotometer (Wilmington, DE) and an Agilent 2100 Bioanalyzer (Santa Clara, CA) were used to determine concentration, purity and integrity of RNA extractions. RNA samples were stored at -20°C until used.

Gene Chip Hybridization

cDNA preparation, and biotin-labeled cRNA generation and fragmentation were performed using Affymetix Genechip 3‘ IVT Express Kits per the manufacturer protocol. Hybridization solution was prepared with the fragmented cRNA and controls, and was then hybridized to the probe array with a 16-hour incubation period. The solution was then removed and the probe array was washed and stained in an automated microfluidics station. Arrays were read by an Affymetrix GeneChip Scanner attached to a workstation running Affymetrix Microarray Suite.

Experimental Design

Gene chip data from five treatments groups were collected. Each treatment had two replicates. Group 1: Stationary phase *S. cerevisiae* cells (grown for seven days in aerated media) washed and resuspended in phosphate buffer saline solution (PBS). This group served as the baseline cultured *S. cerevisiae* control to which the RNA expression levels from sample groups 2-5 were compared. Groups 2-4 were all stationary phase *S. cerevisiae* cells (grown for seven days in aerated media) encapsulated in silica via cell-directed assembly (CDA) and stored dry different lengths of time under ambient temperature and relative humidity (RH). Group 2 was stored for four hours. Group 3 was stored for three days. Group 4 was stored for seven days. Group 5 was stationary phase *S. cerevisiae* cells (grown for seven days in aerated media) encapsulated in silica via cell-directed assembly (CDA) and stored dry for seven days under ambient temperature and relative humidity (RH), then returned to YPD+A broth and incubated at 30°C for four hours before extraction of RNA.

Identification of differentially expressed genes

Differentially expressed genes between 4-hours post-encapsulation (4H) vs control, 3-days post-encapsulation (3D) vs control, 7-days post-encapsulation (7D) vs control, and 7-days post-encapsulation plus 4-hours reincubation period (7D+4HRI) vs control were identified using LIMMA R package.[75] GCRMA was used to normalize the genes expression intensity value.[22] The differentially expressed genes were identified using FDR-adjusted P-value <0.05 and (|log2(fold

change) | > 1.

Functional enrichment analysis

Functional enrichment of differentially expressed genes (FDR-adjusted P-value < 0.05) was performed using the Database for Analysis, Validation, and Integrated Discovery system (DAVID).[26] The Gene Ontology (GO) categories included are biological process (BP), cellular component (CC), and molecular function (MF). The statistical significance of enriched GO categories was estimated by Expression Analysis Systematic Explorer (EASE) scores. To minimize the redundancy between GO categories, clustering of GO terms was carried out using functional annotation tool in DAVID. The geometric mean of EASE score of individual GO categories in each cluster was used to assign overall EASE score to each functional cluster. Functional cluster with Enrichment Score of >2.0 were considered significant using Yeast array 2.0 as a background.

Gene network analysis and Yeast global metabolic pathway

Protein-Protein interaction network of differentially expressed genes ((|log2(fold change) | > 2) shared between 4-hours, 3-days, 7-days post-encapsulation, and 7-days post-encapsulation+4-hours reincubation was carried out using String v10 software.[36] Nodes in the network represents protein products coded by differentially expressed genes while edges represent the physical interaction between the products of different genes. The significance of interaction was assessed with very high confidence score of 0.9.

Differentially expressed genes shared between 4-hours, 3-days, 7-days post-encapsulation, and 7-days post-encapsulation+4-hours reincubation are mapped on KEGG[78] global metabolic map using IPATH.[101] Metabolic compounds are represented by nodes while edges represent the enzymatic alterations.

Results and Discussion

The differentially expressed genes between 4-hours post-encapsulation vs control, 3-days post-encapsulation vs control, 7-days post-encapsulation vs control, and 7-days post-encapsulation plus 4-hours reincubation vs control are summarized in **Figure 3.1**. Overall, 460 (upregulated) and 588 (downregulated) genes are shared between all four gene sets (**Figure 3.1A, 3.1B**). This major

overlap among each of the time points shows that there is a general encapsulation response by cells regardless of the duration of encapsulation, which persists for at least four hours after re-incubation. The second major overlap was differentially upregulated (209) and downregulated (243) genes shared between 7-days post-encapsulation and 7-days post-encapsulation+4hours reincubation (**Figure 3.1A, 3.1B**). This implies that restoration of the pre-encapsulation state within four hours of re-incubation is far from complete. The yeast suffers from a genomic version of post-traumatic stress syndrome. The third major overall commonality was differentially upregulated (113) and downregulated (131) genes shared between 4-hours and 3-days post-encapsulation (**Figure 3.1A, 3.1B**). However, 84 (upregulated) and 123 (downregulated) genes are uniquely differentially expressed at 7-days post-encapsulation+4-hours reincubation followed by 69 (upregulated) and 85 (downregulated) genes are uniquely differentially expressed at 4-hours post-encapsulation. These two sets may be considered the unique or special responses to re-incubation and encapsulation, respectively.

Functionally enriched GO clusters

The largest subgroup of genes (upregulated: 466, downregulated: 588) are those shared among all the treatment categories (4H, 3D, 7D, 7D+4HRI). **Figure 3.2** illustrates the functionally enriched GO categories among these genes. The upregulated genes in this group are most enriched with a GO cluster of terms related to membrane. A second cluster is enriched with terms related to sporulation and ascospore cell wall assembly (**Figure 3.2A**). This is consistent with the finding that stress of Saccharomyces cells result in sporulation.[86] Furthermore, the upregulated genes in this group are enriched with GO cluster of terms related to cell wall organization. Again, this is consistent with published findings that cells use cell wall integrity pathways (CWI) to maintain cell wall and reorganization of cell wall when subjected to stress.[89,100]

The downregulated genes shared between 4H, 3D, 7D and 7D+4HRI are enriched with GO cluster of terms related to ribosome, translation, cytoplasm translation and ribonucleoprotein complex (**Figure 3.2C**). Downregulation of all these categories is implicated in published studies of cellular stress response.[3,100] The normal growing cell contains ~200,000 ribosomes[3]; thus biogenesis of ribosomes require ample amount of energy, and downregulation of these processes is expected to conserve cellular energy.[3] Furthermore, downregulated gene are enriched in TCA cycle (**Figure**

3.2C). Activity in the TCA cycle leads to higher amount of reactive oxygen species (ROS) that can result in cell death in the stressed cell.[102] Downregulation of the TCA cycle is characteristic of persister cells.[102]

The subgroup with the second most genes (upregulated: 209, downregulated: 243) are shared between 7-days post-encapsulation and 7-days post-encapsulation+4-hours reincubation. This set of genes represents a genomic "memory" of encapsulation that persists for at least four hours after reincubation in growth medium. The upregulated component includes clusters related to fungal type cell wall organization, response to stress, membrane, and integral component of membrane (**Figure 3.3A**). The functional GO clusters enriched in differentially downregulated genes shared between 7-days post-encapsulation and 7-days post-encapsulation+4-hours reincubation include clusters related to protein degradation such as proteasomal complex, proteolysis, ATP binding and ATP transport, amino acid transport, transmembrane transport, and antiporter activity (**Figure 3.3B**).

The third largest group of genes (upregulated: 113 (**Figure 3.2A**), downregulated: 131 (**Figure 3.4**)) are those shared between 4-hours post encapsulation and 3-days post encapsulation. GO clusters enriched in differentially expressed genes shared between 4-hours post-encapsulation and 3-days post-encapsulation are shown in figure 4. The upregulated genes are ontologically random; there are no clusters at our threshold value. However, the downregulated genes show statistical enrichment for one cluster, with subclusters related to ribosome, ribosome biogenesis, rRNA processing, cytoplasmic translation, and a second GO cluster related to protein transport (**Figure 3.4**).

The fourth largest group of genes (upregulated: 84, downregulated: 123) are expressed only at 7-days post-encapsulation+4-hours reincubation (7DR+4HRI). Differentially upregulated genes are enriched in three clusters of GO categories related to membrane, fungal type vacuole and protein transport (**Figure 3.5A**). Differentially downregulated genes are enriched with single cluster of GO categories related to GTPase activity (**Figure 3.6C**). This set represents the manner in which the post-incubation was uniquely different from both the pre-encapsulation cells and the during-encapsulation cells.

The fifth largest group of genes (upregulated: 69, downregulated: 85) uniquely differentially expressed at 4-hours post-encapsulation (4H) are illustrated in **Figure 3.7**. The GO enriched clusters uniquely upregulated at 4-hours post-encapsulation includes integral component of membrane and plasma membrane (**Figure 3.7A**). Ion transport, especially iron ion transport and iron ion homeostasis categories, are enriched in upregulated genes uniquely expressed at 4-hours post-encapsulation (**Figure 3.7A**). Iron ion transport has been implicated in stress response as an important element for growth of a cell.[103,104] The GO Cluster related to vacuole, such as membrane vacuole, is also enriched in upregulated genes (**Figure 3.7A**). Vacuoles are membrane bound organelles that play a vital role in cellular ion homeostasis.[105] Yeast cells growing in nutrient enriched medium transport the cytosolic iron, beyond that required for metabolic processes, to vacuoles. Under iron-poor environment, stored iron in vacuole is transported back to the cytosol.[105] The enrichment of upregulated genes in iron transport and vacuole transport support this observation of iron transport to cytosol through membrane vacuole under stressed environment (**Figure 3.7A**).

The GO clusters enriched in genes uniquely downregulated at 4-hours post-encapsulation include clusters of ribosome, translation, ribosome biogenesis, rRNA processing (**Figure 3.7C**). All these processes are implicated under environmental stress response and play vital role in adjusting cell's metabolic state in response to stress.[2,3,100] Other groups on the Venn diagram contain only a few genes each. In some cases, they show no enriched categories, while in other cases the judgement of enrichment is based on only a very small number of genes.

Comparison of gene expression patterns of encapsulated cells with universal yeast stress set

Gasch et al[2,3] analyzed the gene expression patterns of *Saccharomyces cerevisiae* in response to diverse environmental conditions such as temperature shocks, amino acid starvation, hydrogen peroxide, hypo and hyper osmatic stress etc. Authors identified ~600 genes whose expression was reduced and ~300 genes whose expression were induced in response to most of these environmental conditions and termed these genes as Environmental Stress Response (ESR) set. We carried out the functional enrichment analysis of ESR set using the same methods our

experimental sets, and compared the results with functional enrichment of encapsulate cells reported in this study. Due to large number of functionally enriched GO clusters in ESR, we only reported the top 5 clusters (**Table 3.1, Table 3.2**). The upregulated ESR genes are functionally enriched with Clusters of GO terms related to oxidation-reduction process, carbohydrate metabolic process, response to oxidative stress, and glutathione metabolic process (**Table 3.1**). The downregulated ESR genes are enriched with Clusters GO terms related to ribosome biogenesis, rRNA processing, rRNA methylation, cytosolic translation, and methylation (**Table 3.2**).

The functional enrichment of downregulated ESR set is in almost complete agreement with the functional enrichment of downregulated genes in encapsulated cells. Results of both these studies shows regardless of nature of stress, cell usually preferred to downregulate/reduce the energy consuming process such translation, transcription, ribosome biogenesis, and rRNA to minimal level as a general response to stress. However, functional enrichment of upregulated ESR set is different from the functional enrichment of encapsulated cells. This is consistent with the published results showing that the upregulated gene categories in response to different types of stress are different from each other, while the downregulated categories are similar.[2,3] Encapsulation stress is different yet from all the others; that is, functional enrichment of upregulated genes in encapsulated cells are mostly enriched with membrane, plasma membrane and cell wall organization related GO terms (**Figure 3.2, 3.3 and Figure 3.5, 3.6**). The closest analogy in nature to the silica confinement is in biofilms, in which the cells are confined in a polysaccharide matrix.[10,94,106] This natural encapsulation also induces upregulation of genes related to cell wall and cell membrane, as well as downregulation characteristic of our experiments and the ESR set.[2,3,11,94,107]

Global metabolic map and protein-protein interaction network

The relationship between the protein products of differentially expressed genes ($|\log_2(\text{fold change})| > 2$) shared between 4H, 3D, 7D, and 7D+4HRI is illustrated using String 10.1[36] (**Figure 3.8A**). The protein-protein interaction network of these genes shows a high degree of variability among the genes in their degree of interactivity. Genes that have more than 5 interactions are shown in a histogram (**Figure 3.8B**). The histogram of genes having >5 interaction is dominated by downregulated gene (green) than upregulated genes. For example, GUA1 (>20 interactions)

encodes GMP synthase that is negatively regulated under nutrients limited conditions. Furthermore, many highly interacting genes (> 5 interactions) are dominated by genes whose products are involved in ribosomal biogenesis, rRNA processing such as such as RRP12, RPS9B, RPL13A, LSG1, RPL2A *etc.* and downregulated at all encapsulation time points. Previous microarray studies shown that genes encoding these transcripts are tightly co-regulated[2,3] which is reinforced by our results and the representation in Figure 8B. However, there are only 7 upregulated genes whose products have > 5 interactions such ALD3 gene produce Aldehyde Dehydrogenase whose expression is induced under stress, CDC40 is involved in cell cycle regulation, VPS71 plays role in protein sorting, and MRPS18 encodes mitochondrial ribosomal protein and vital for cell's viability.

Conclusions

This study was designed to explore what changed and what remained constant upon and during encapsulation of yeast cells in a silica matrix, and upon subsequent reincubation in a growth medium. One overall result is that, for most genes (3,180 out of 5,900 on the gene chip), the expression levels did not change significantly from the initial stationary state in culture. Of the remaining 2,720 differentially expressed genes the expression level of a large number (1,048 differentially expressed gene), changed upon encapsulation and thereafter remained relatively constant throughout 7 days of encapsulation and even upon reincubation in growth medium. Gene Ontology analysis revealed that, throughout the course of our experiments, the categories of downregulated genes agreed with categories seen in other experiments on yeast under stress from various treatments, namely ribosome, ribosome biogenesis, rRNA processing, cytoplasmic translation, and translation.[3] On the other hand, our upregulated categories, especially related to cell wall organization and cell membrane, were unique to the type of stress our experiments imposed. A transient change in expression of genes at 4-hours post-encapsulation iron ion transport, that was not observed later during the encapsulation at 3-days, 7-days, and 7-days post-encapsulation+4-hours reincubation, showed strong upregulation of iron ion transport. We attribute this to facilitation of iron transport from vacuoles to cytoplasm.

The downregulated gene expression patterns observed in this study such as reduction in ribosome biogenesis and minimal transcription and translation, are typical features of a quiescent/dormant

or persister state. On the other hand, the persister state in nature, for example in biofilms, is characterized by the ability to return to the state that existed prior to the stress that induced persistence. Our experiments are inconclusive in that regard. The four hours reincubation after release from encapsulation may not have provided enough time for the culture to move through a growth phase and return to the stationary state that pertained prior to encapsulation. It is reasonable to hypothesize that our cells, given sufficient time, after an initial growth phase, would have returned to that stationary state. We therefore believe that our encapsulation induced a persister state, with downregulated categories of genes in common with other stressed or persister states, and upregulated gene categories specific to the type of stress induced by the silica matrix encapsulation.

This work has implications for use of yeast cells embedded in silica matrices for devices or sensors. It suggests that such use will depend on advances on two fronts. One such necessary advance would be to somehow embed the cells in a less stressful way. From our previous study[100] it may be that aqueous gel encapsulation may be less stressful, Time-varying studies on that method should be carried out. The second advance would be to design the desired function of the cell in such a way as to be compatible with the constraints of the persister state.

Figures

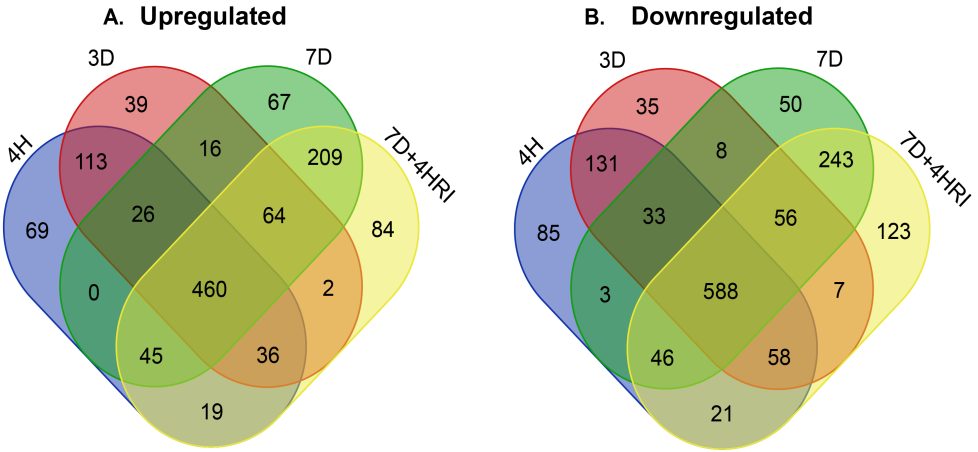


Figure 3.1. Distribution of all differentially expressed genes (FDR-adjusted P-value <0.05) that overlap between 4-hours post-encapsulation(4H), 3-days post-encapsulation (3D), 7-days post-encapsulation (7D), and 7-days post-encapsulation+4hours re-incubation (7D+4HRI).

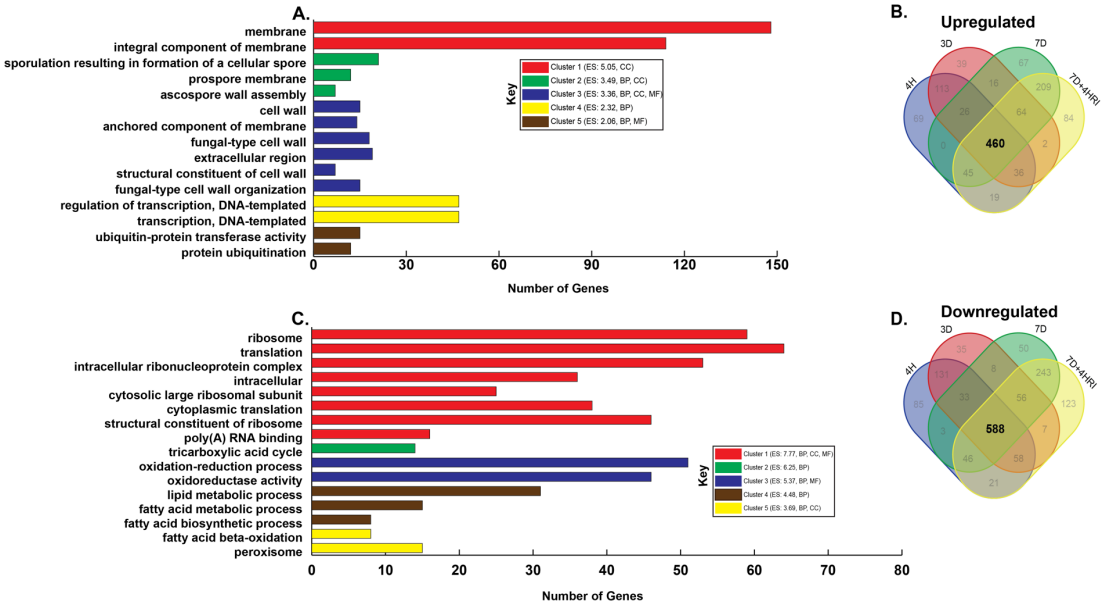


Figure 3.2. Functionally enriched GO clusters shared among 4H, 3D, 7D, and 7D+4HRI that have enrichment score > 2 and corresponding Gene Ontology (GO) biological process (BP), cellular component and molecular function (MF) categories. (A) GO clusters enriched in upregulated genes shared between 4H, 3D, 7D, and 7D+4HRI shown in (B). (C) GO clusters enriched in downregulated genes shared between 4H, 3D, 7D, and 7D+4HRI shown in (D).

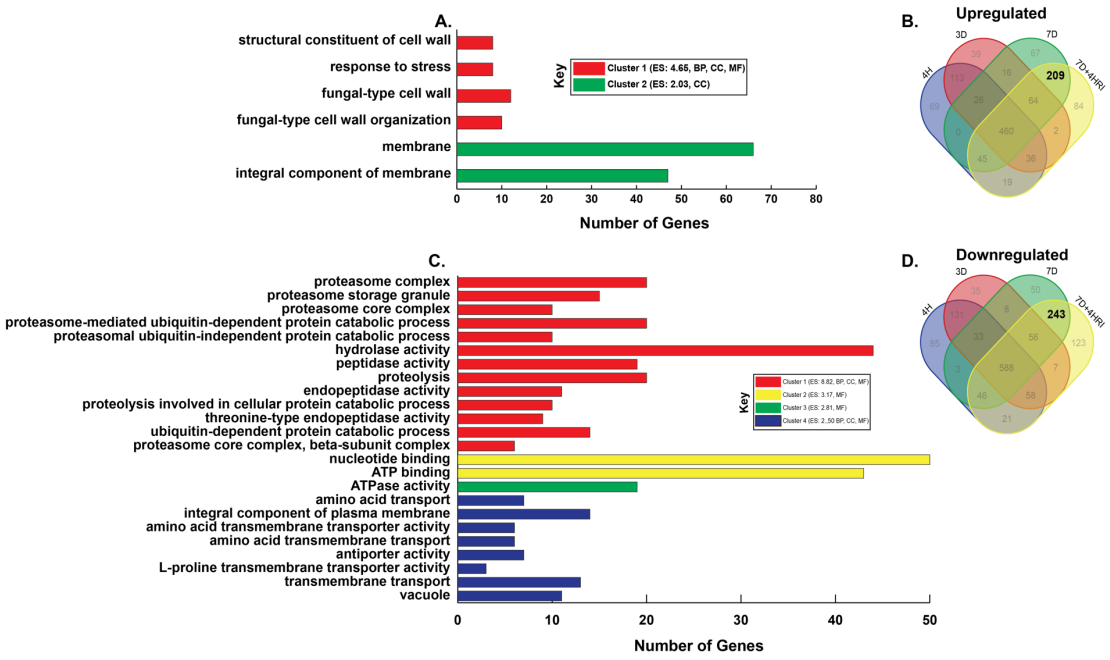


Figure 3.3. Functionally enriched GO clusters shared among 7D, and 7D+4HRI that have enrichment score > 2 and corresponding Gene Ontology (GO) biological process (BP), cellular component and molecular function (MF) categories. (A) GO clusters enriched in upregulated genes shared between 7D, and 7D+4HRI shown in (B). (C) GO clusters enriched in downregulated genes shared between 7D, and 7D+4HRI shown in (D).

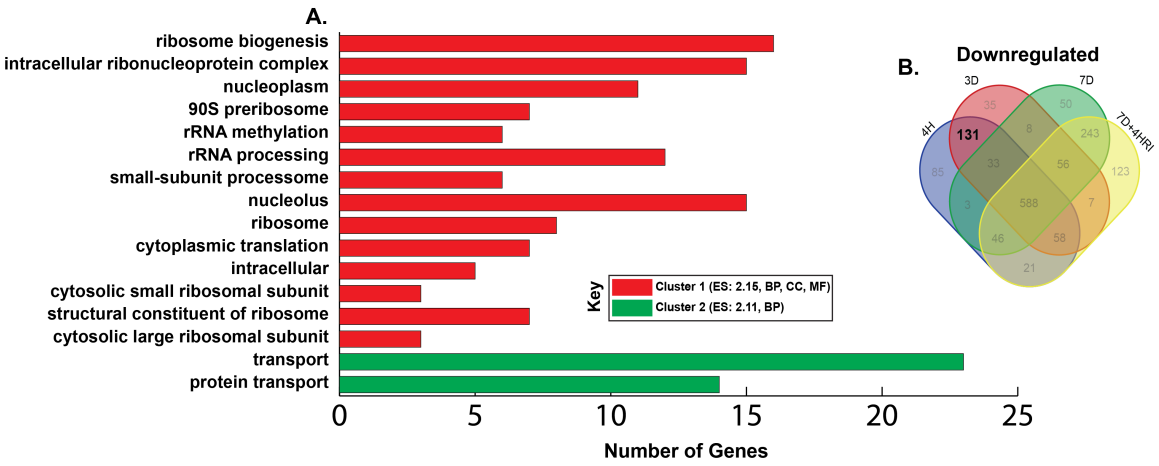


Figure 3.4. Functionally enriched GO clusters shared among 4H and 3D that have enrichment score > 2 and corresponding Gene Ontology (GO) biological process (BP), cellular component and molecular function (MF) categories. (A) GO clusters enriched in downregulated genes shared between 4H and 3D shown in (B).

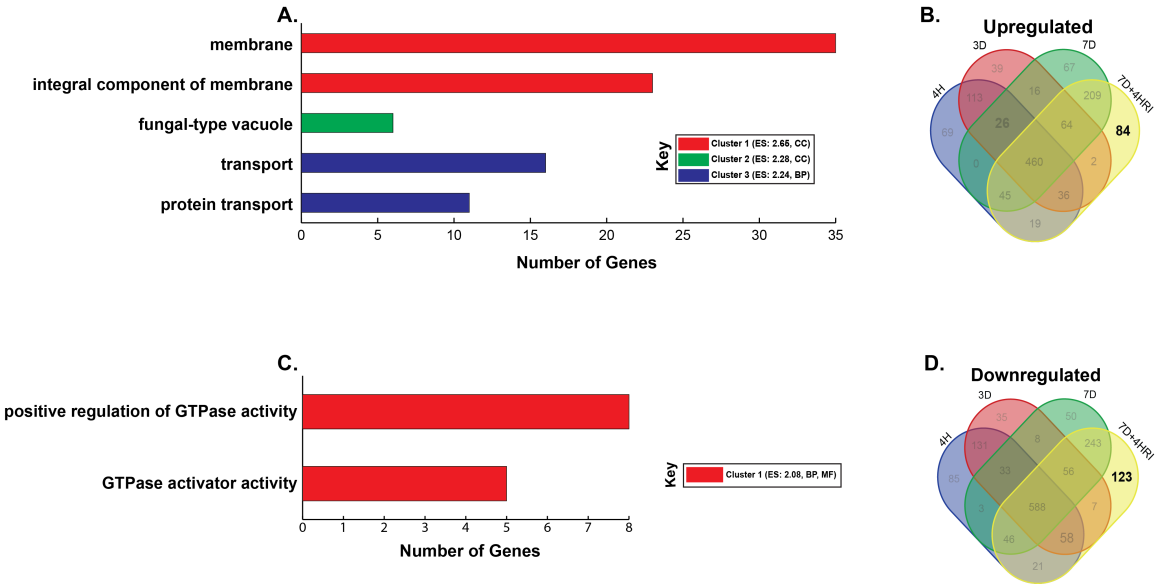


Figure 3.5. Functionally enriched GO clusters unique in 7D+4HRI that have enrichment score > 2 and corresponding Gene Ontology (GO) biological process (BP), cellular component and molecular function (MF) categories. (A) GO clusters enriched in upregulated genes uniquely expressed at 7-days post-encapsulation+4-hours reincubation (7D+4HRI) as shown in (B). (C) GO clusters enriched in downregulated genes uniquely expressed at 4hours post-encapsulation (4H) as shown in (D).

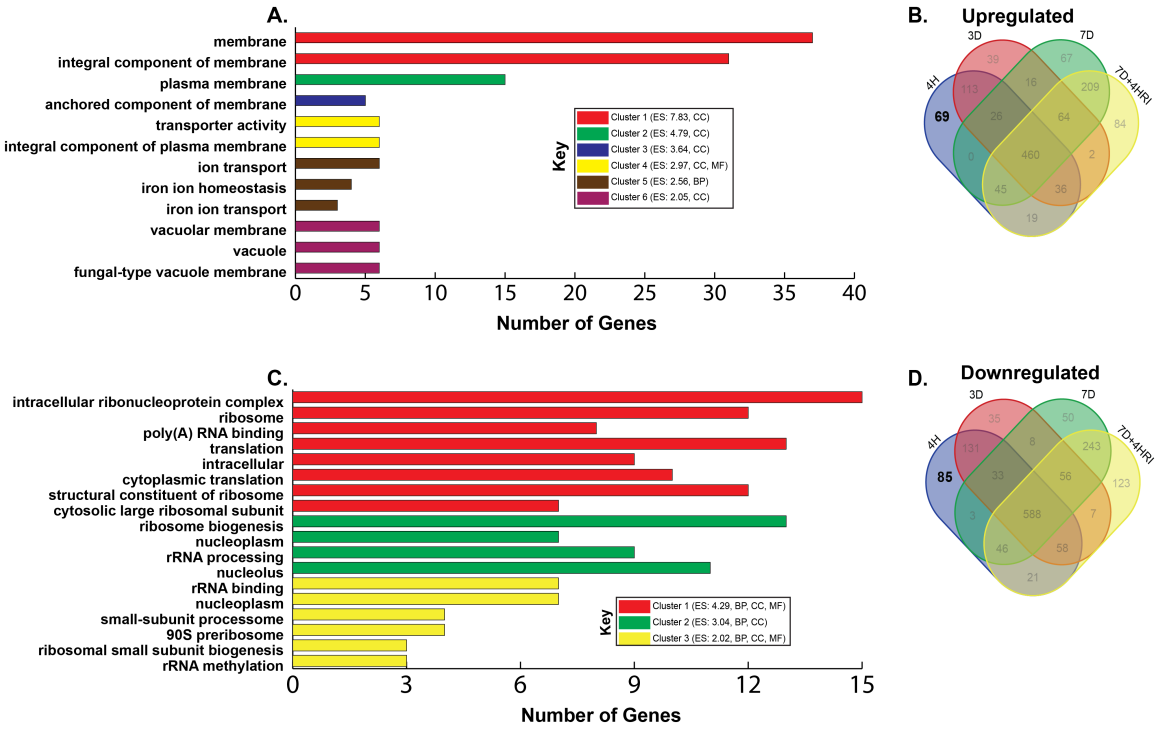


Figure 3.6. Functionally enriched GO clusters unique in 4H that have enrichment score > 2 and corresponding Gene Ontology (GO) biological process (BP), cellular component and molecular function (MF) categories. (A) GO clusters enriched in upregulated genes uniquely expressed at 4-hours post-encapsulation (4H) as shown in (B). (C) GO clusters enriched in downregulated genes uniquely expressed at 4-hours post-encapsulation (4H) as shown in (D).

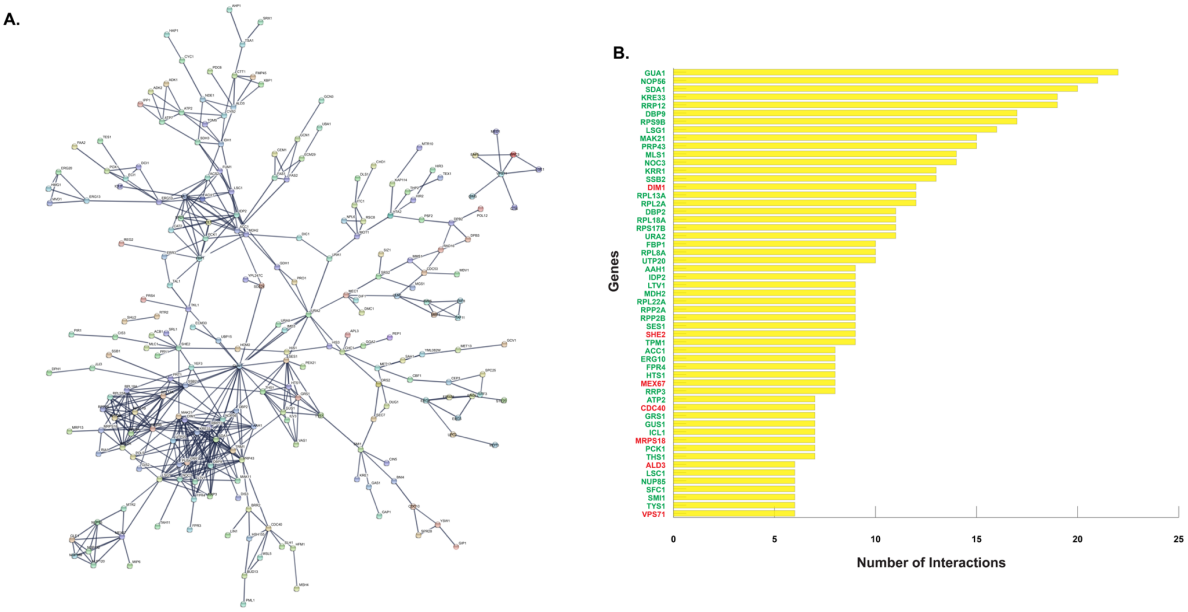


Figure 3.7. (A) Network of differentially expressed genes shared between 4H, 3D, 7D, and 7D+4HRI. (|log2(fold change) | > 2) visualized using STRING with very high confidence score of 0.9. (B) The histogram of genes from (A) having >5 interaction. Red color represent upregulated genes and Green color shows downregulated genes.

Tables

Table 3.1 Functionally enriched GO clusters in ESR induced genes (Top 5)

Cluster	Category	Enrichment Score	GO Terms
Cluster1	BP, MF	10.21	oxidation-reduction process, oxidoreductase activity
Cluster2	BP	5.59	carbohydrate metabolic process
Cluster3	BP, MF	5.11	cellular oxidant detoxification, glutathione peroxidase activity
Cluster4	BP, MF	3.35	peroxidase activity, response to oxidative stress, peroxiredoxin activity
Cluster5	BP, MF	3.27	glutathione metabolic process, glutathione transferase activity

Table 3.2 Functionally enriched GO clusters in ESR repressed genes (Top 5)

Cluster	Category	Enrichment score	GO Terms
Cluster1	BP, CC	101.08	ribosome biogenesis, nucleolus, rRNA processing
Cluster2	CC	51.85	nucleolus, nucleus
Cluster3	BP, CC, MF	41.56	intracellular ribonucleoprotein complex, cytoplasmic translation, translation, ribosome, cytosolic large ribosomal subunit, structural constituent of ribosome, intracellular, cytosolic small ribosomal subunit
Cluster4	BP, CC	31.04	90S preribosome, rRNA methylation, small-subunit processome, endonucleolytic cleavage in ITS1 to separate SSU-rRNA from 5.8S rRNA and LSU-rRNA from tricistronic rRNA transcript, endonucleolytic cleavage in 5'-ETS of tricistronic rRNA transcript, endonucleolytic cleavage to generate mature 5'-end of SSU-rRNA
Cluster5	BP, MF	6.77	methyltransferase activity, methylation, transferase activity

Chapter 4: Using Optimal F-Measure and Random Resampling in Gene Ontology Enrichment Calculations[2]

Abstract

Ordinarily Gene Ontology (GO) enrichment analysis is subject to an arbitrary threshold for defining significance of enriched classes. In this paper, we consider replacing an arbitrary threshold with F-measure optimization to define the *p*-value that divides "significant enrichment" from "non-significant". It is found that evaluation of false negatives (essential for computing recall and thus F-measure) requires a heuristic (but reasonable) assumption. We apply F-measure optimization to two sets of genes from different organisms and use Benjamini-Hochberg and random resampling to evaluate the number of false positives. It is found that the uncorrected *p*-value that produces optimum F-measure varies widely from one data set to another. It is also found that all three methods of FDR calculation diverge from each other within a range of uncorrected *p*-values that provide F-measure optimum *p*-values. This study includes in Supplementary Material a pipeline for using resampling and F-measure optimization to create lists of enriched GO classes that provide for variable weights of precision and recall.

Introduction

Gene Ontology (GO) enrichment analysis is a powerful tool to interpret the biological implications of selected groups of genes. The gene lists from experiments such as microarrays, are gathered into clusters associated with biological attributes, and defined as GO terms.[30] The GO terms are arranged in an acyclic tree structure from more specific to more general descriptions, including biological process (BP), cellular component (CC), and molecular function (MF).[30] GO aspires to be both a cross-species common language, and means of understanding the uniqueness of each species at in the categories or biological process, location in the cell, and molecular function.[30] Each enriched GO term is then evaluated by its significance level, i.e. the probability that the enrichment has not occurred by pure chance.

Enrichment tools have been developed to process large gene lists to generate significantly enriched

[2] This work has been submitted in BMC Bioinformatics for publication.

ontologies. Huang *et.al* (2009) summarizes the tools widely used for GO enrichment.[108] Different tools emphasize different features. Gorilla[109], DAVID[26], g-profiler[110] are web interfaces that integrate functional annotations including GO annotations, disease and pathway databases etc. Blast2GO[111] extends annotation of gene list to non-model organisms by sequence similarity. GO-Miner[112], Babelomics[24], FatiGO[113], GSEA[25], and ErmineJ[114] apply resampling or permutation algorithms on random sets to evaluate the number of false positives in computed gene ontologies associated with test sets. DAVID[26] and Babelomics[24] introduced level-specific enrichment analysis; that is, not including both parents and children terms. The TopGO algorithms "eliminate" and "parent-child" eliminate or reduce the weight of genes in the enriched children terms when calculating parent term enrichment.[115] TopGO[116] and GOstats[117] provide R-scripted tools for ease of further implementation. Cytoscape plugin in BinGO[118] is associated with output tree graphs.
To calculate raw *p*-values for GO enrichment without multiple hypothesis correction, methods used include hypergeometric distribution, Fisher's distribution, Binomial distribution, or χ^2 distribution.[119] Rivals *et. al*. discussed the relative merits of these methods.[119]

Uncorrected *p*-values are subjected to multiple hypothesis correction by the methods of Bonferroni[32], Benjamini-Hochberg (BH)[31], or Benjamini-Yekutieli (BY).[120] Bonferroni is the most stringent one among these multiple hypothesis correction methods. Benjamini-Hochberg has been widely applied in enrichment tools such as BinGO[118], DAVID[26], GOEAST[121], Gorilla[109], and Babelomics[24], to name a few. The Benjamini-Yekutieli method is included in the GOEAST package.[121] GOssip provides a direct analytical estimation of false positives that compares well with resampling.[122] In random resampling, a null set is constructed by random sampling from the same structured database that the test set enrichment is computed from. Because it is most directly related to the question of how likely it is that an enrichment result may arise by chance, it can be reasonably considered the most reliable method for estimating false positives.[122] Resampling is more computer-intensive than other methods[122], but high-throughput techniques have demonstrated that it is possible to keep resampling time in a reasonable range.[25]

In applying all the cited methods and tools, it is common to apply a threshold boundary between "significant enrichment" and "insignificance". Such assignment to one of two classes is an example of a binary classification problem. Often such classifications are made utilizing an optimum F-

measure.[123] Rhee, *et.al.* (2008) have suggested application of F-measure optimization to the issue of gene ontology enrichment analysis.[124] In the present work, we present an approach to gene enrichment analysis based on F-measure optimization, which considers both precision and recall and provides a flexible reasonable threshold for data sets depending on user choice as to the relative importance of precision and recall. The work suggests that resampling is preferable to analytical methods to estimate FDR, and F-measure optimization is preferable to an arbitrary threshold, in computing enrichment in gene ontology analysis.

Methods

Enrichment Tool

For results reported in this study (described below), the TopGO[116] package is implemented to perform GO enrichment analysis, using the "classic" option. In this option, the hypergeometric test is applied to the input gene list to calculate an uncorrected *p*-value.

FDR Calculation

The empirical resampling, and Benjamini-Hochberg (BH) methods are used to estimate the FDR. The *p*-value adjustment using Benjamini-Hochberg corrections are carried out by a function implemented in the R library. http://stat.ethz.ch/R-manual/R-devel/library/stats/html/p.adjust.html The resampling method is based on the definition of *p*-value as the probability that an observed level of enrichment might arise purely by chance. To evaluate this probability, we generate several null sets, which are the same size as the test set. The genes in the null sets are randomly sampled from the background/reference list. GO enrichment analysis was carried out on both test set and null set. The average number of enriched results in the null sets would be the false positives. In all the results shown in this paper, 100 null sets were used to compute the average. In the pipeline, available for download in Supplementary material, the number of null sets is an adjustable parameter. The ratio of false positives to total positives is the FDR.

F-measure Optimization

The F-measure is a weighed value of precision and recall.[123] The parameter β is chosen based on the research question, whether minimization of "type I" (false positive) or "type II" (false negative) error, or balance between the two, is preferred, according to the equation:

$$F_\beta = (1+\beta^2)\frac{Precision \cdot Recall}{\beta^2 Precision + Recall} \quad \text{---------- } Equation\ 1$$

The larger the magnitude of β the more the value of F_β is weighted towards recall; the smaller the value of β the more the value of F_β is weighted towards precision. To obtain F-measure, precision and recall are derived from enrichment results. For an analytical method such as BH, we first calculate precision by the equation: $Precision = 1 - FDR$. The true positive is derived by $(True\ Positive) = (Total\ Positive) \cdot Precision$. On the other hand, for the resampling method, the number of enriched terms from random set indicates false positives. Therefore, we first calculate true positive number by:

$True\ Positve = (Total\ Positive) - (False\ Positive)$.

Then, we calculate the precision:

$$Precision = \frac{True\ Positive}{Total\ Positive}$$

Recall is defined as

$$Recall = \frac{True\ Positve}{Relevant\ Elements}$$

"Relevant Elements" is defined by

$$Relevant\ Elements = True\ Positives + False\ Negatives$$

In the absence of the ability to calculate "False Negatives" directly, we estimate the number of relevant elements as the maximum true positive achieved across the range of possible *p*-values, as shown graphically in **Figure 4.1** for BH method of computing false positives, for a gene list to be described in detail later in the paper. In this figure we plot total positives, false positives (False Discovery Rate x total positives, and true positives (total positives – false positives) vs. uncorrected *p*-value for the entire range of *p*-values from 0 to 1. At very relax *p*-values the FDR approaches 1, resulting in the true positives approaching 0. It is difficult to evaluate false negatives and thus assign a number for "relevant elements", since a false negative is an object that escaped observation, and thus can't be counted directly. Yet such estimation is essential to applying F-measure. In our case, if we follow the trajectory of the true positives in **Figure 4.1** as the threshold is relaxed, we see that at very stringent *p*-values all positives are true positives. As the threshold is relaxed further, more false positives are generated, so the total positive and true positive curves start to diverge. At $p = 0.2$ (a far higher value than would ordinarily be used as a cutoff) the true

positives reach a maximum, and the number of true positives starts to decline as *p* is further relaxed. We utilize this maximum value as the maximum number of GO categories that can be possibly regarded as enriched in the data set; i.e., the number of relevant elements.

Based on precision and recall at each raw *p*-value cut-off, we can obtain a table and curve of F-measure vs raw *p*-value. Optimizing F-measure provides us a threshold which emphasize precision ($\beta<1$) or recall ($\beta>1$), or balance of both ($\beta=1$). Note that precision and recall are extreme values of F-measure; that is, Precision=F_0 and Recall=F_∞.

Data Sets

Environmental Stress Response (ESR)

First dataset is the Yeast Environmental Stress Response (ESR) data[3], a robust data set for a model organism. The ESR set is list of genes commonly differentially expressed in response to environmental stresses such as heat shock, nutrient depletion, chemical stress, etc. Approximately 300 genes are up-regulated and 600 genes are down-regulated in ESR set.[3] We expect this set is "well-behaved" (give reasonable results with standard methods of analysis), since the data come from a very well annotated model organism subject to a widely studied experimental intervention.

Alarm Pheromone (AP)

The second data set is comprised of human orthologs to the honey bee Alarm Pheromone set.[125] The Alarm Pheromone set is a list of genes differentially expressed in honey bee brain in response to the chemical alarm pheromone, which is a component of the language by which honey bees communicate with each other. Previous studies have shown that the Alarm Pheromone set is enriched in placental mammal orthologs, compared to other metazoans including non-social insect orthologs.[126] The Alarm Pheromone set is much smaller than the ESR set, with 91 up-regulated genes and 81 down-regulated genes. We expect the AP set is not so "well-behaved" compared to the ESR set, as we are using a model organism orthologs (human) to a non-model organism (honey bee) and the organisms diverged about 600 million years ago.

Random Test Sets

To generate a baseline of the analysis for each data set using different FDR calculation methods,

we have applied the pipeline to analyze randomly-generated sets as "test" set inputs, where FDR should equal to 1 for all uncorrected *p*-values.

The BH FDR curves are calculated in the following way: The R program p.adjust is applied to generate a list of analytically calculated FDR's (BH) corresponding to uncorrected *p*-values for each "test" sets. Then the lists of FDRs are merged and sorted by uncorrected *p*-values. The FDRs are smoothed by a "sliding window" method: at each uncorrected *p*-value point, the new FDR is the average value of 11 FDRs centered by the uncorrected *p*-value point.

The Resampling FDR curves are calculated in the following way: The output uncorrected *p*-values are binned in steps of 1E-4. The counts in each bin for the "test" set enrichment categories are the "Total Positives", and average counts in each bin for the null set enrichment categories are the "False Positives". The process is repeated for the 10 "test" sets within each group, and we would obtain 10 lists of "Total Positives" and "False Positives". Then the number of "Total Positives" and "False Positives" are averaged. The FDR would be the quotient of the averaged Total and False Positives. Then, all the FDRs are plotted against the uncorrected *p*-values.

Results and Discussion

In this section, we present the results of applying our methods to the two previously published sets of data introduced in the Methods section, the ESR set and the human orthologs of the Alarm Pheromone set. For both above data sets, we show the results from analyzing the genes using the biological process (BP) category of the gene ontology.

ESR Set (Environmental Stress Response, yeast)

Benjamini-Hochberg (BH)

Figure 4.2 shows the results of F-measure optimization on the ESR data based on FDR calculated by Benjamini-Hochberg (BH) method. As expected by their definitions, precision (F_0) decreases with increasing *p*-value while recall increases with increasing *p*-value. $F_{0.5}$ (precision-emphasized), F_1 (precision and recall equally weighted) and F_2 (recall-emphasized) all show relative maxima, providing a rational basis for assigning a threshold for significance. The horizontal scale is extended far enough to visualize the determination of the number of relevant items. In the case of the up-regulated gene set, maximum F_1 occurs at an uncorrected *p*-value close

to 0.05. In the case of the down-regulated gene set however, it appears that a more stringent cutoff would be appropriate.

Resampling

Figure 4.3 shows the results of F-measure optimization on the ESR data using resampling to calculate FDR. The false positives are calculated by average number of GO categories enriched in random sets. All the F-measures optimize at much lower uncorrected *p*-values than do the F-measures calculated by the BH method.

Alarm Pheromone Set (human orthologs)

Benjamini-Hochberg (BH)

Figure 4.4 shows exactly the corresponding results as **Figure 4.2**, this time on the human orthologs to the honey bee alarm pheromone set. F-measures are maximized at much higher thresholds than for the ESR set. The difference in optimal F-measure is largely due to the different shapes of the recall curves. For the ESR set, precision drops significantly more rapidly with increasing uncorrected *p*-value than does the AP set. Therefore, a higher uncorrected *p*-value can be used for the latter set with essentially the same degree of confidence.

Resampling

Figure **4.5** shows the number of GO categories and F-measures for the alarm pheromone set human orthologs using resampling method. The resampling method have found more false positives than BH, and therefore the precision is much lower than the precision calculated from BH, and the F-measures are optimized at lower uncorrected *p*-values than the F-measures calculated from BH.

In the above **Figures 4.2-4.5**, we can note the stepped structure in the number of enriched GO categories. The stepped structure lies in the fact that the number of genes associated with any GO category, in the test set or reference set, must be an integer with limited number of choices. Therefore, the uncorrected *p*-values calculated would be in a discrete set instead of a continuum. Consequently, the number of positives as a function of *p*-values increases in a stepped way. As a result, the F-measures derived from the number of GO categories have spikes. But as our graphs have demonstrated, the optimal F-measures reflect the different weights on precision and recall despite the spikes.

Comparison of FDR (False Positive) Calculation by Benjamini-Hochberg (BH) and Resampling

In the previous section, we have demonstrated how to use F-measure optimization to obtain a flexible threshold based on requirement of the research problem, whether precision or recall is more heavily weighted. This section demonstrates how the resampling method applies to the F-measure optimization approach by providing an alternative way to estimate FDR. We have plotted the FDR calculated by BH and Resampling of the test sets, as well as the randomly selected sets that are same size as the test sets as the baselines. The random "test" sets are selected from the same reference set as the test sets they are compared to. To reduce random fluctuation, each random "test" set result is averaged over 50 runs.

Figure 8 shows that for the ESR set, the BH method and resampling estimate similar FDR at low *p*-value. As the threshold increases, the BH method estimates lower false discovery rate, and therefore higher precision, than the resampling method at the same raw *p*-value. For the Alarm Pheromone set, the BH method estimates lower FDR than resampling.

To further evaluate the methods, we carried out multiple runs using random sets as test sets. In this case, the FDR should in principle be 1, for any uncorrected *p*-value. The results of this test are shown in **Figure 4.6a**, where for each segment of *p*-values (bin size = 0.0001) we show the mean plus/minus the standard deviation. The Resampling method passes the test on the average, but the results are noisy; and the BH method systematically underestimates FDR. **Figure 4.6b** shows that the noise in the Resampling method results in **Figure 4.6a** are largely due to the variation in the random "test" sets, and that the noise level in using random resampling for real data is acceptably low.

Threshold comparison summary

In this section, we show the bar graphs (**Figures 4.7, 4.8**) of number of enriched biological process GO categories associated with ESR and alarm pheromone sets, at different thresholds including the commonly-used FDR < 0.05, optimization of $F_{0.5}$, F_1, and F_2, with BH and resampling methods. This is an alternative representation of data already presented in **Figures 4.2-4.5**.

In **Figures 4.7 and 4.8**, we are comparing the number of enriched GO categories found using

thresholds calculated by BH and Resampling. The leftmost bars in each cluster represents FDR under 0.05, and the next three bars are results from flexible thresholds by F-measure optimization. We see that the most widely used method (BH FDR<0.05) is generally quite stringent. When we weight more on precision by optimizing $F_{0.5}$ using the resampling method, a more stringent threshold is calculated for the alarm pheromone set. Maximizing F_1 would bring back many more terms. Investigation into precision tells us that sometimes the F_1-maximized FDR is too high (near 1) for us to tell apart signal and noise, and might not be a good threshold. However, whether F_1 optimization is reasonable depends on the data set. In the ESR set, where precision and recall can reach the balance where both are reasonable (precision=0.84, recall =0.96 for the up-regulated; precision=0.93, recall=0.94 for the down-regulated), the F_1 optimized threshold is reliable in the sense we can be confident in validity of the large majority of the terms that are returned. On the other hand, for the alarm pheromone set, precision becomes very low when F_1 is optimized (precision=0.46, recall=0.35 for the up-regulated; precision=0.53, recall=0.81 for the down regulated), so the user may wish to use a more stringent threshold. The major point is that the threshold will not be arbitrary, but rather based on the scientist's judgment on the relative biological significance of how much weight to give precision and recall.

Conclusions

In this work, we have addressed two issues with the commonly used methods in the GO enrichment analysis: the arbitrariness of the threshold for significance, and the relationship between resampling vs. Benjamini-Hochberg theory for estimating false discovery rate. For the first part, we introduced optimization of F-measures so that both type I and II errors are considered. Unlike arbitrarily applied threshold of BH FDR<0.05 or raw *p*-value<0.01 for any data set, the F-measure optimization approach provides a flexible threshold appropriate to the nature of the data set and the research question. If the data set is high in noise-to-signal ratio and the penalty for letting in false positive is high, we can choose to optimize F-measures weighing more on precision. If the data set fails to show much enrichment by commonly-applied methods, we can relax the threshold and extract the best information indicated by F-measure optimization. For the second part, we introduced resampling approach using random sets to directly estimate false positives, and consequently derive values of FDR, precision, recall, and F-measures. We believe that for the GO enrichment analysis, a resampling method is more universally applicable than the BH method,

because the resampling method is a direct algorithmic representation of the false discovery rate; that is, the likelihood of getting a positive result by pure chance. Thus, results from resampling are independent of the structure of the data set, such as parent-child relationships. A concern is that, because of the nature of the problem, we were forced to use a heuristic (albeit reasonable) method to estimate the false negatives, essential for calculating recall. We judge that this concern is more than offset by the advantage of enabling the replacement of an arbitrary threshold with F-measure optimization. In the supplementary material, we present our automatic pipeline integrating TopGO with resampling and analyzing functions to carry out the whole process of resampling, enrichment analysis, F-measure calculation, and representing results in tables and figures. The pipeline also includes a GOstats[117] module for easy analysis of under-represented terms. As demonstrated, the pipeline can also calculate analytical FDR including, but not limited to the BH method. In summary, we suggest replacing a fixed *p*-value for assigning a threshold in enrichment calculations with an optimal F-measure, which incorporates the well-established and well-defined concepts of precision and recall.

Figures

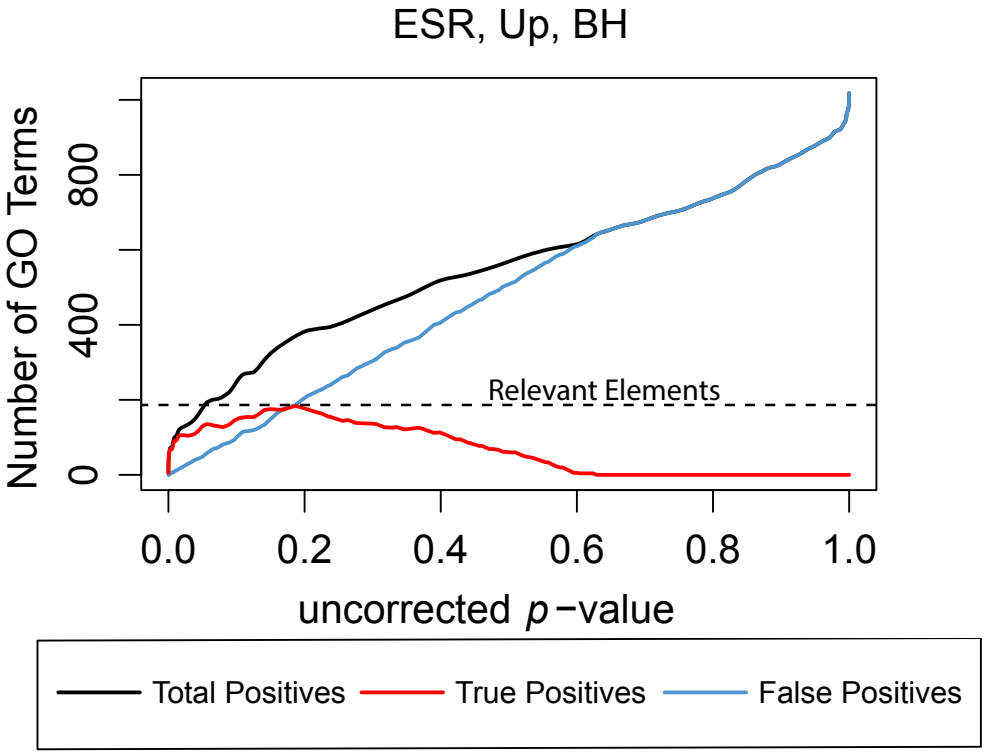


Figure 4.1. Number of positives for the yeast environmental stress response (ESR) set over the full range of uncorrected p-values from 0 to 1. "Total Positives" is the number of Biological Process GO categories returned as a function of the *p*-value threshold for significance. "False Positives" is the number of total positives multiplied by the False Discovery Rate as calculated by the Benjamini-Hochberg formulation. "True Positives" is "Total Positives" minus "False Positives". "Relevant Items", necessary to estimate number of false negatives, is estimated as the largest number of true positives computed at any uncorrected *p*-value.

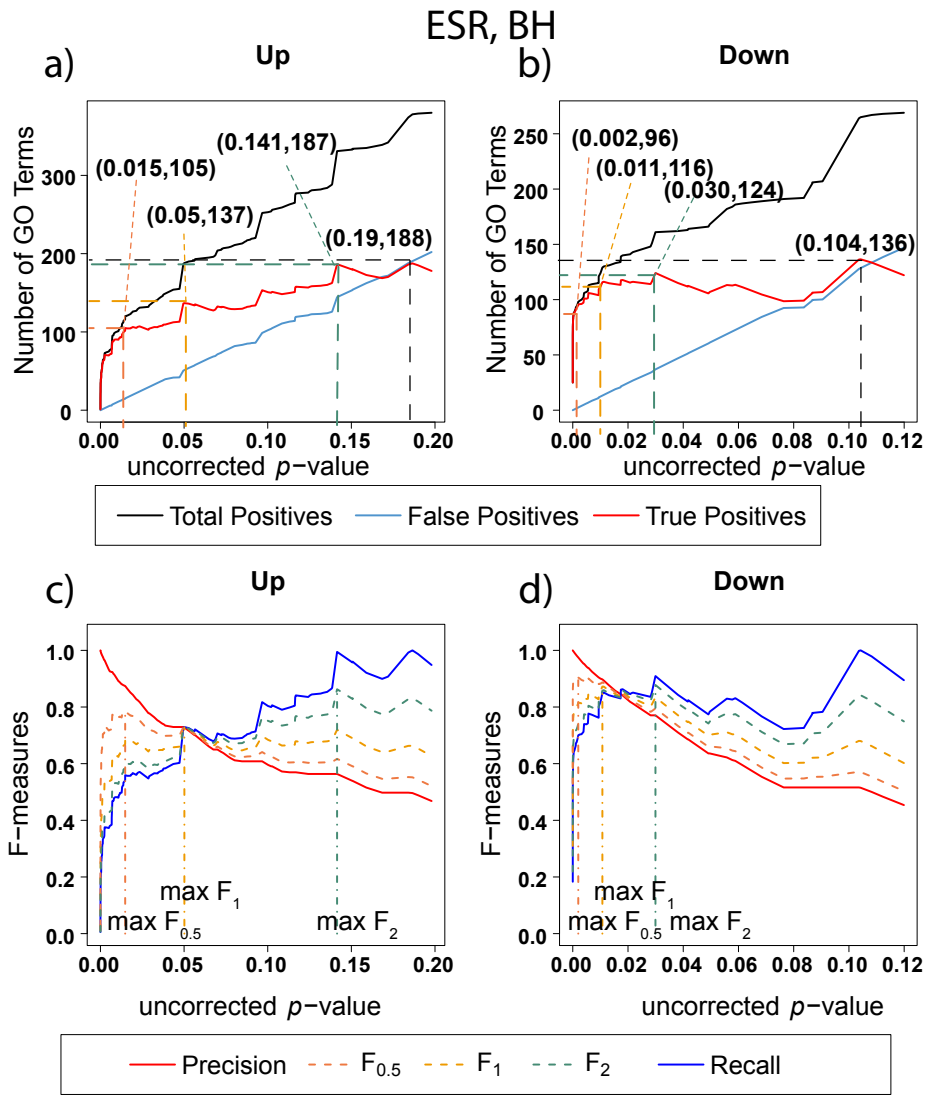


Figure 4.2. Number of positives and F-measure values for ESR set, BH-estimated FDR a) Shows the number of enriched biological process Gene Ontology categories as a function of uncorrected *p*-value, the Benjamini-Hochberg number of false discoveries, and the projected true positives, namely the difference between the total positives and the false positives, for the upregulated ESR gene set. This panel is from the same data set at **Figure 4.1**. The number pairs in parenthesis are respectively (uncorrected *p*-value maximizing $F_{0.5}$, "number of true positives" at that *p*-value), (uncorrected *p*-value maximizing F_1, "number of true positives" at that *p*-value), (uncorrected *p*-value maximizing F_2, "number of true positives" at that *p*-value), (uncorrected *p*-value maximizing true positives, "number of true positives" at that *p*-value) b) is the same as a) for the downregulated gene set. c) shows the F-measures computed from a) and d) the F-measures computed from b). Number of relevant items, necessary to calculate recall (and therefore (F-measure)), is approximated by (total positives – false positives) $_{max}$, where F_{max} is the *p*-value at which the computed true positives are a maximum. This *p*-value is 0.19 for upregulated gene list (a) and at 0.104 for downregulated gene list. (b) The pairs of numbers in parenthesis in a) and b) indicate the *p*-value and number of returned GO terms at significant markers, specifically at maximum $F_{0,5}$

Figure 4.2 (Con't)
(emphasizing precision), F_1 (balanced emphasis between precision and recall), F_2 (emphasizing recall), and F_{max} (*p*-value at which computed true positives are a maximum).

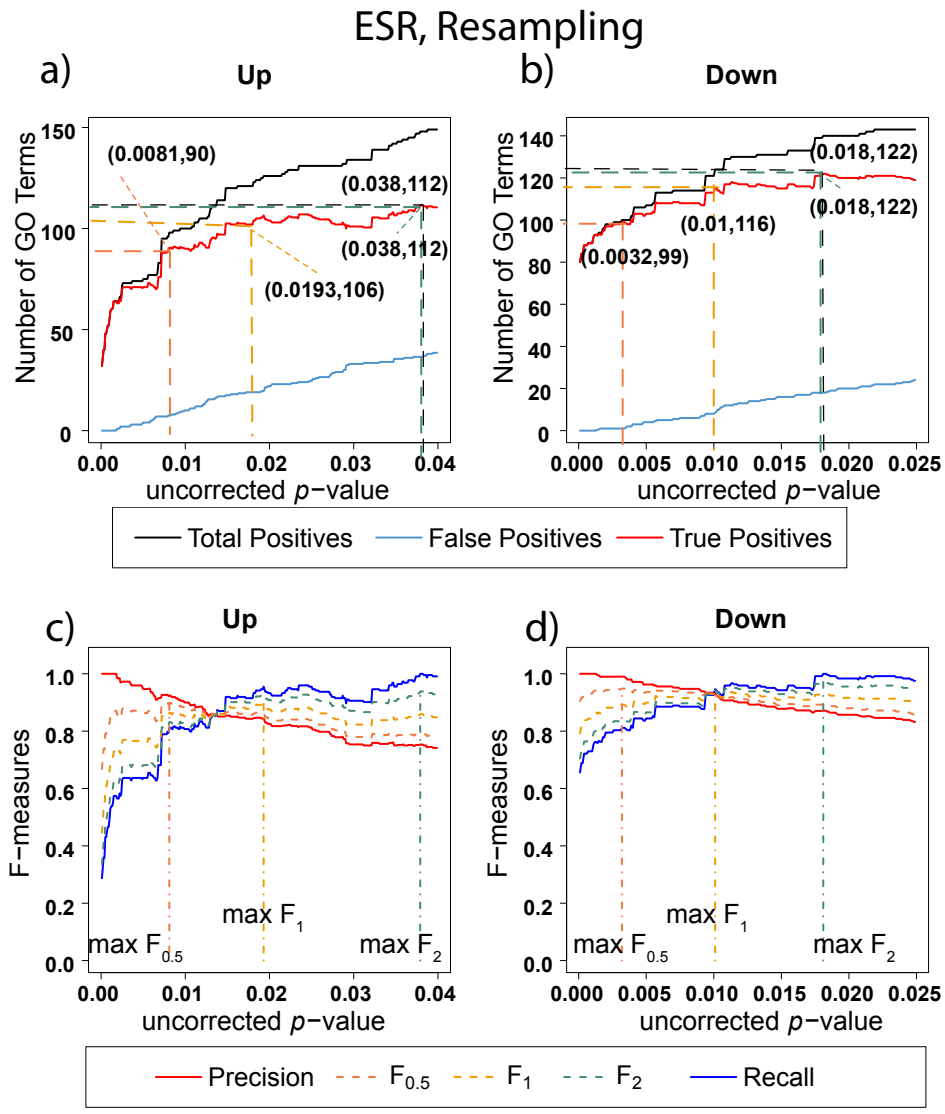


Figure 4.3. Number of Positives and F-measure values for ESR set, Resampling-estimated FDR. A) Shows the number of enriched biological process Gene Ontology categories as a function of uncorrected *p*-value, the average number of enriched Gene ontology categories from the random set as the false positives, and the projected true positives, namely the difference between the total positives and the false positives, for the up-regulated ESR gene set. The number pairs in parenthesis are respectively (uncorrected *p*-value maximizing $F_{0.5}$, "number of true positives" at that *p*-value), (uncorrected *p*-value maximizing F_1, "number of true positives" at that *p*-value), (uncorrected *p*-value maximizing F_2, "number of true positives" at that *p*-value), (uncorrected *p*-value maximizing true positives, "number of true positives" at that *p*-value) b) is the same as a) for the down-regulated gene set. c) shows the F-measures computed from a) and d) shows the F-measures computed from b). Number of relevant items, necessary to calculate recall (and therefore (F-measure)), is approximated by (total positives – false positives) $_{max}$, where F_{max} is the *p*-value at which the computed true positives are a maximum. This *p*-value is 0.038 for upregulated gene list (a) and 0.018 for downregulated gene list. (b) The pairs of numbers in parenthesis in a) and b) indicate the *p*-value and number of returned GO terms at significant markers, specifically at

Figure 4.3 (Con't)
maximum $F_{0,5}$ (emphasizing precision), F_1 (balanced emphasis between precision and recall), F_2 (emphasizing recall), and F_{max} (*p*-value at which computed true positives are a maximum).

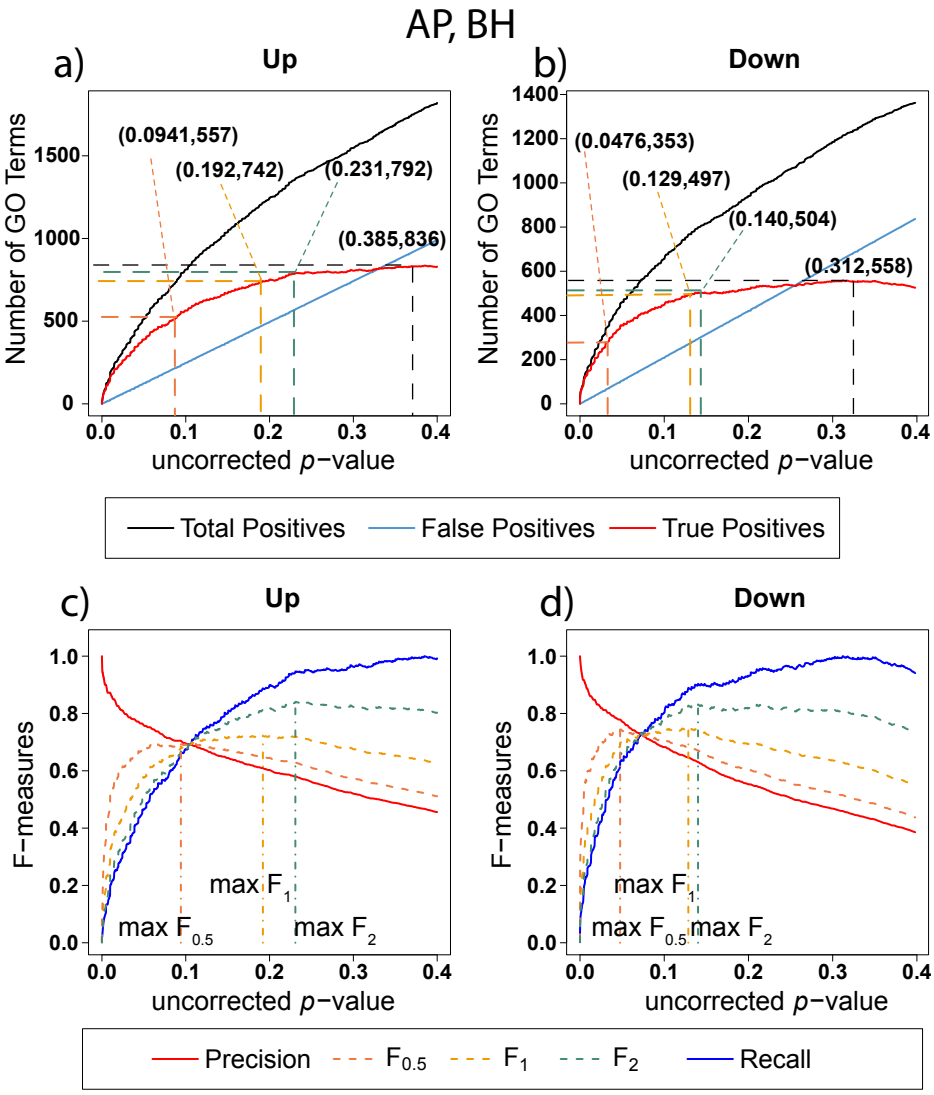


Figure 4.4. Number of Positives and F-measure values for Alarm Pheromone set, BH-estimated FDR a) Shows the number of enriched biological process Gene Ontology categories as a function of uncorrected *p*-value, the Benjamini-Hochberg number of false discoveries, and the projected true positives, namely the difference between the total positives and the false positives, for the upregulated alarm pheromone human orthologs gene set. The number pairs in parenthesis are respectively (uncorrected *p*-value maximizing $F_{0.5}$, "number of true positives" at that *p*-value), (uncorrected *p*-value maximizing F_1, "number of true positives" at that *p*-value), (uncorrected *p*-value maximizing F_2, "number of true positives" at that *p*-value), (uncorrected *p*-value maximizing true positives, "number of true positives" at that *p*-value) b) is the same as a) for the downregulated gene set. c) shows the F-measures computed from a) and d) the F-measures computed from b). Number of relevant items, necessary to calculate recall (and therefore (F-measure)), is approximated by (total positives – false positives) $_{max}$, where F_{max} is the *p*-value at which the computed true positives are a maximum. This *p*-value is 0.385 for upregulated gene list (a) and at 0.312 for downregulated gene list. (b) The pairs of numbers in parenthesis in a) and b) indicate the *p*-value and number of returned GO terms at significant markers, specifically at maximum $F_{0,5}$

Figure 4.4 (Con't)
(emphasizing precision), F_1 (balanced emphasis between precision and recall), F_2 (emphasizing recall), and F_{max} (*p*-value at which computed true positives are a maximum).

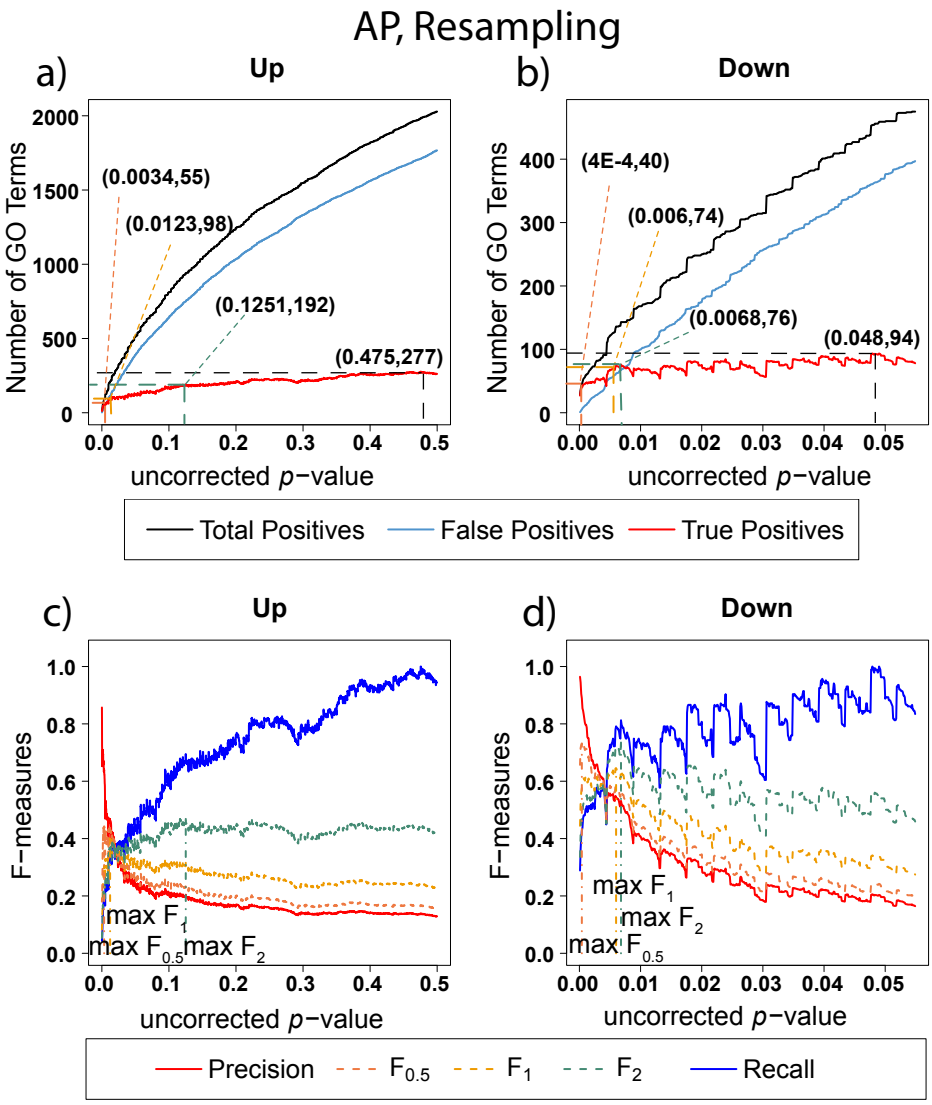


Figure 4.5. Number of Positives and F-measure values for Alarm Pheromone set, Resampling-estimated FDR. The Figure shows the number of enriched biological process Gene Ontology categories as a function of uncorrected *p*-value, the average number of enriched Gene ontology categories from the random set as the false positives, and the projected true positives, namely the difference between the total positives and the false positives, for the up-regulated alarm pheromone human orthologs gene set. b) is the same as a) for the down-regulated gene set. c) shows the F-measures computed from a) and d) the F-measures computed from b). Number of relevant items, necessary to calculate recall (and therefore (F-measure)), is approximated by (total positives – false positives) $_{max}$, where F_{max} is the *p*-value at which the computed true positives are a maximum. This *p*-value is 0.475 for upregulated gene list (a) and at 0.048 for downregulated gene list. (b) The pairs of numbers in parenthesis in a) and b) indicate the *p*-value and number of returned GO terms at significant markers, specifically at maximum $F_{0,5}$ (emphasizing precision), F_1 (balanced emphasis between precision and recall), F_2 (emphasizing recall), and F_{max} (*p*-value at which computed true positives are a maximum).

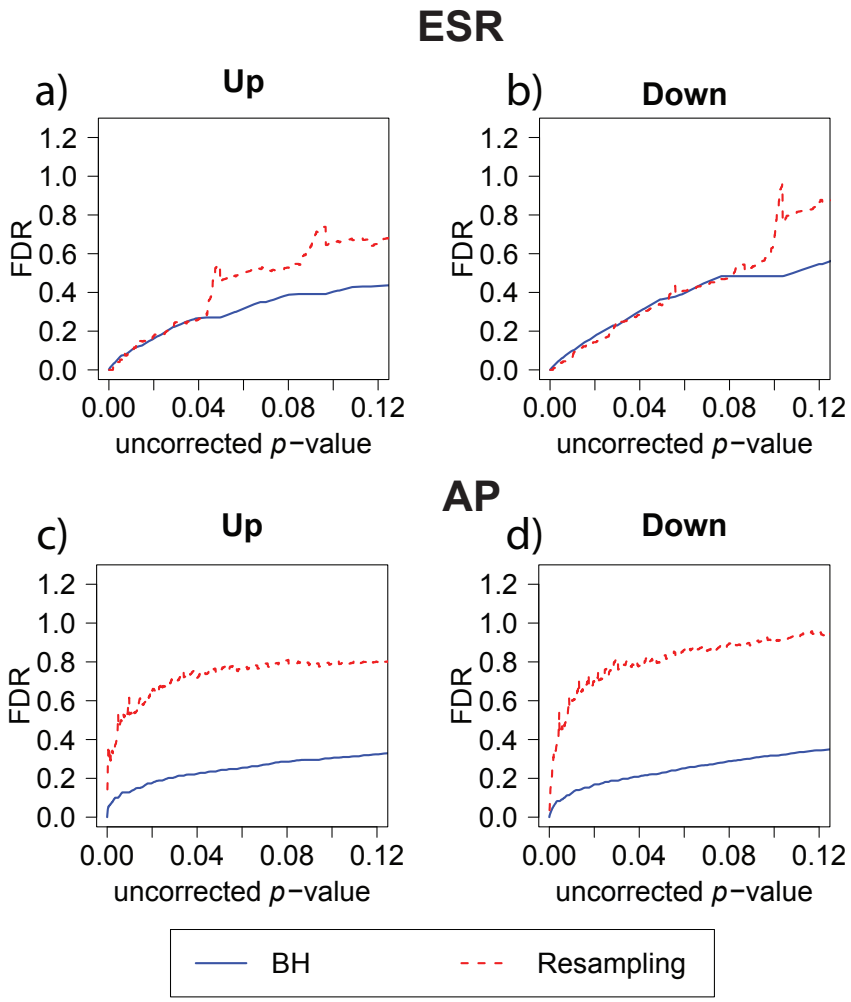


Figure 4.6. False discovery rate estimated by Benjamini-Hochberg and Resampling for the ESR set and Alarm Pheromone set. **Figure 4.8** compares the number of false discovery rate calculated by Benjamini-Hochberg and Resampling in each set: a) up-regulated ESR, b) down-regulated ESR, c) up-regulated Alarm Pheromone set, and d) down-regulated Alarm Pheromone set. Generally, resampling has found higher false discovery rate than Benjamini-Hochberg. At low *p*-values, the BH and resampling methods get similar estimation of false discovery rate for the ESR set.

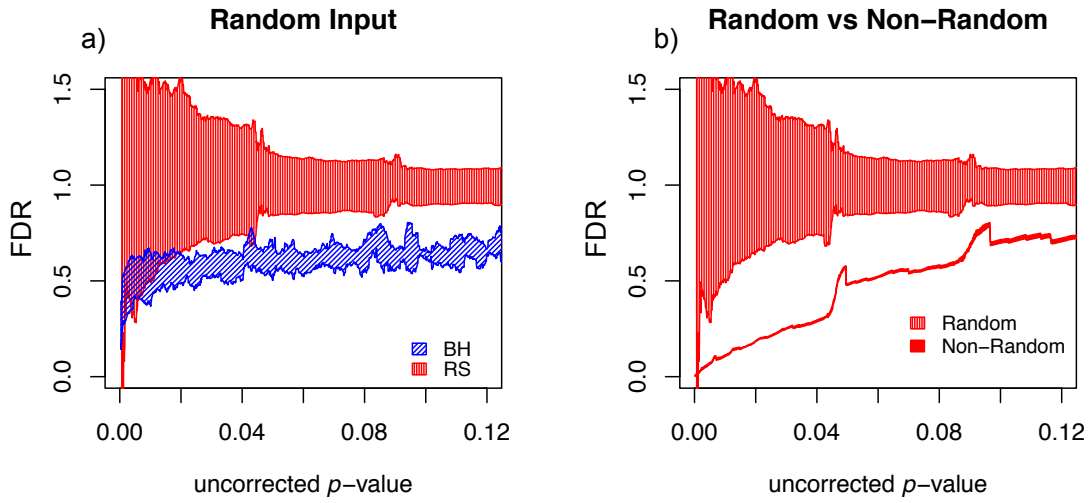


Figure 4.7. Comparison of different FDR calculation methods on accuracy and convergence. a) Comparison of BH and Resampling on random "test" sets. At each *p*-value (*p*-values binned at intervals of .0001), the mean and standard deviation are calculated and plotted as shown. The random test sets consist of 281 yeast genes, against the background of the entire yeast genome. For each of the methods 50 test sets were used and the mean plus/minus standard deviation plotted as shown, it is seen that BY hits the mark (FDR=1) over a wide range of *p*-values, Resampling hits the mark on the average but with substantial noise, while BH systematically underestimates FDR. b) Evaluation of resampling convergence on a real data set, ESR upregulated considered in this paper. This set is run against five different ensembles of null sets, each ensemble containing 100 null sets. The mean and standard deviation are plotted and compared to the results from the random test sets. It is seen that the noise of the resampling method on a real data set is acceptable.

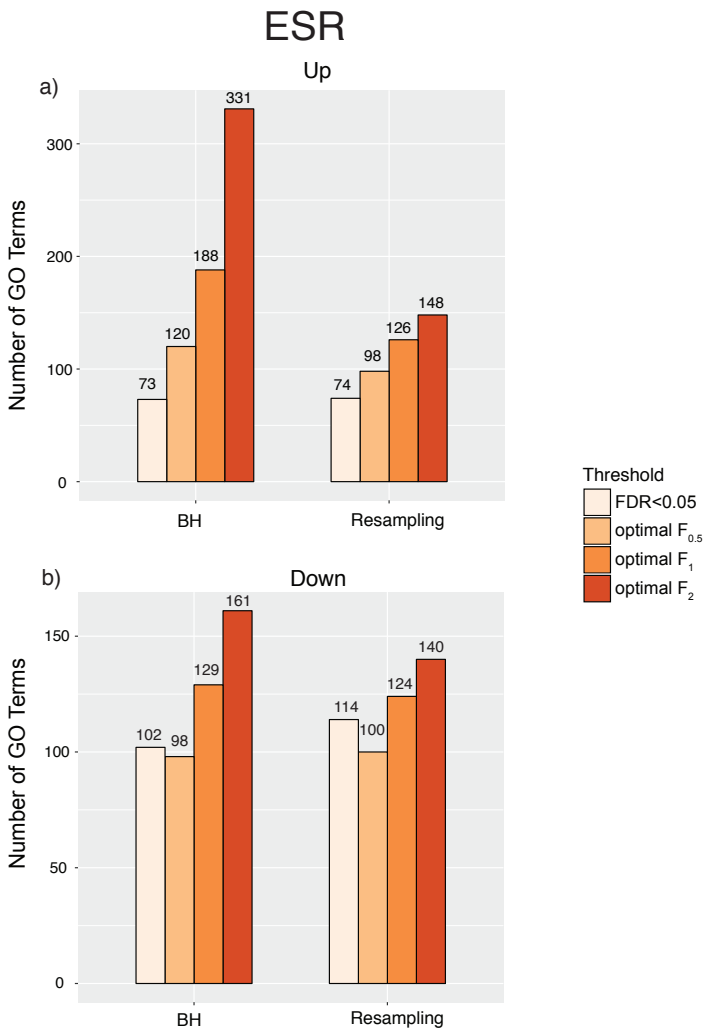


Figure 4.8. Number of GO categories obtained at different thresholds including: FDR<0.05, $F_{0.5}$ optimization, F_1 optimization, and F_2 optimization, using Benjamini-Hochberg (BH) and Resampling. Calculated FDR for the ESR set, up- and down-regulated. BH-estimated $F_{0.5}$, F_1, and F_2 optimization gives 120, 188, and 331 GO terms for up-regulated ESR set respectively, and 98, 129, 161 terms for down-regulated set respectively. Resampling-estimated $F_{0.5}$, F_1, and F_2 optimization gives 98, 126, and 148 terms for up-regulated ESR set respectively, and 100, 124, 140 terms for down-regulated set respectively. As more emphasis is placed on recall, the threshold would increase and more terms would be recovered. The F-measure optimization thresholds estimated by BH is more relaxed by resampling and consistently brings back more terms. The FDR<0.05 threshold is more stringent for the up-regulated set than $F_{0.5}$ optimization, but less stringent for the down-regulated set than $F_{0.5}$ optimization. For both up- down- regulated sets, the most stringent thresholds given by BH and resampling are close to each other (FDR<0.05 for up-regulated, and $F_{0.5}$ optimization for down-regulated), indicating BH and resampling estimates similar FDR at the most stringent thresholds for the ESR set, but then deviates as thresholds increase.

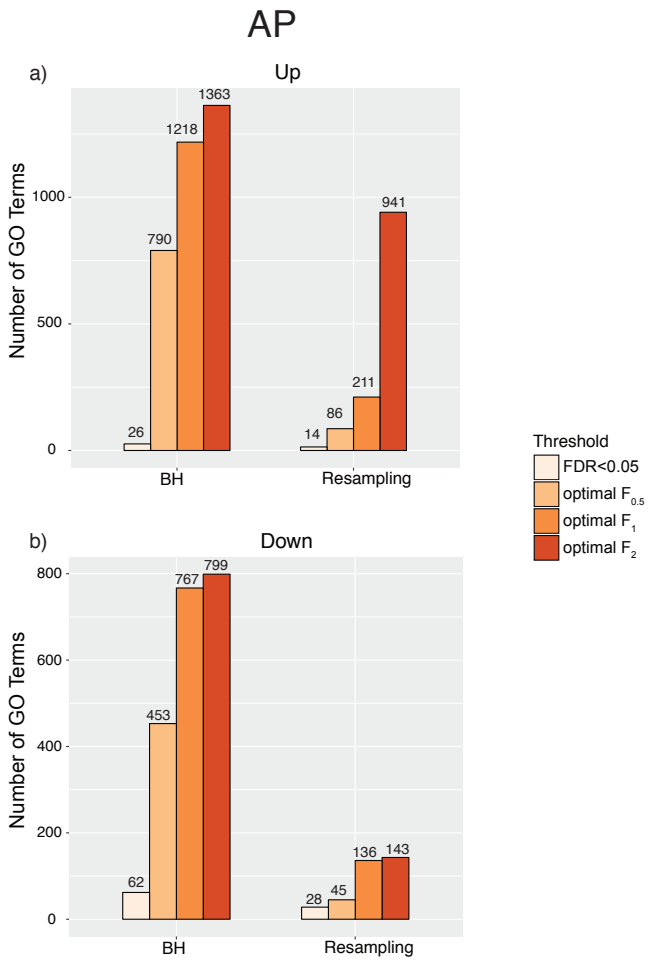


Figure 4.9. Number of GO categories obtained at different thresholds including: FDR<0.05, $F_{0.5}$ optimization, F_1 optimization, and F_2 optimization, using Benjamini-Hochberg (BH) and Resampling – calculated FDR for the AP set, up- and down-regulated. BH-estimated $F_{0.5}$, F_1, and F_2 optimization gives 790, 1218, and1363 GO terms for up-regulated AP set respectively, and 453, 767, 799 terms for down-regulated set respectively. Resampling-estimated $F_{0.5}$, F_1, and F_2 optimization gives 86, 211, and 941 terms for up-regulated AP set respectively, and 45, 136, 143 terms for down-regulated set respectively. As more emphasis is placed on recall, the threshold would increase and more terms would be recovered. The F-measure optimization thresholds estimated by BH is more relaxed by resampling and consistently brings back more terms. The FDR<0.05 threshold is more stringent for the up-regulated set than $F_{0.5}$ optimization, but less stringent for the down-regulated set than $F_{0.5}$ optimization. Increasing the thresholds to optimize $F_{0.5}$, an F-measure which includes the effect recall but still emphasize on precision, brings in many more terms. Therefore, for the alarm pheromone set a cutoff of FDR<0.05 might leave out too many possible candidates.

References

(1) Goffeau, A.; Barrell, B. G.; Bussey, H.; Davis, R. W.; Dujon, B.; Feldmann, H.; Galibert, F.; Hoheisel, J. D.; Jacq, C.; Johnston, M.; *et al.* Life with 6000 Genes. *Science (80-.).* **1996**, *274*, 546–567.

(2) Gasch, A. P.; Werner-Washburne, M. The Genomics of Yeast Responses to Environmental Stress and Starvation. *Funct. {&} Integr. Genomics* **2002**, *2*, 181–192.

(3) Gasch, A. P.; Spellman, P. T.; Kao, C. M.; Carmel-Harel, O.; Eisen, M. B.; Storz, G.; Botstein, D.; Brown, P. O. Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Mol. Biol. Cell* **2000**, *11*, 4241–4257.

(4) Bigger, J. TREATMENT OF STAPHYLOCOCCAL INFECTIONS WITH PENICILLIN BY INTERMITTENT STERILISATION. *Lancet* **1944**, *244*, 497–500.

(5) Dawson, C. C.; Intapa, C.; Jabra-Rizk, M. A. "Persisters": Survival at the Cellular Level. *PLOS Pathog.* **2011**, *7*, 1–3.

(6) Lewis, K. Persister Cells, Dormancy and Infectious Disease. *Nat. Rev. Microbiol.* **2007**, *5*, 48–56.

(7) Lewis, K. Persister Cells. *Annu. Rev. Microbiol.* **2010**, *64*, 357–372.

(8) Gerdes, K.; Semsey, S. Microbiology: Pumping Persisters. *Nature* **2016**, *534*, 41–42.

(9) Conlon, B. P.; Rowe, S. E.; Gandt, A. B.; Nuxoll, A. S.; Donegan, N. P.; Zalis, E. A.; Clair, G.; Adkins, J. N.; Cheung, A. L.; Lewis, K. Persister Formation in Staphylococcus Aureus Is Associated with ATP Depletion. *Nat. Microbiol.* **2016**, *1*, 16051.

(10) Reynolds, T. B.; Fink, G. R. Bakers' Yeast, a Model for Fungal Biofilm Formation. *Science* **2001**, *291*, 878–881.

(11) Fanning, S.; Mitchell, A. P. Fungal Biofilms. *PLOS Pathog.* **2012**, *8*, 1–4.

(12) Zhu, Y. Immobilized Cell Fermentation for Production of Chemicals and Fuels. In *Bioprocessing for Value-Added Products from Renewable Resources*; 2007; pp. 373–396.

(13) Karel, S. F.; Libicki, S. B.; Robertson, C. R. The Immobilization of Whole Cells: Engineering Principles. *Chem. Eng. Sci.* **1985**, *40*, 1321–1354.

(14) Rathore, S.; Desai, P. M.; Liew, C. V.; Chan, L. W.; Heng, P. W. S. Microencapsulation of Microbial Cells. *Journal of Food Engineering*, 2013, *116*, 369–381.

(15) Vos, P. De; Faas, M. M.; Spasojevic, M.; Sikkema, J.; De Vos, P. Encapsulation for Preservation of Functionality and Targeted Delivery of Bioactive Food components.(Report). *Int. Dairy J.* **2010**, *20*, 292.

(16) Brun-Graeppi, A. K. A. S.; Richard, C.; Bessodes, M.; Scherman, D.; Merten, O. W. Cell Microcarriers and Microcapsules of Stimuli-Responsive Polymers. *Journal of Controlled Release*, 2011, *149*, 209–224.

(17) Johnson, P. E.; Muttil, P.; MacKenzie, D.; Carnes, E. C.; Pelowitz, J.; Mara, N. A.; Mook, W. M.; Jett, S. D.; Dunphy, D. R.; Timmins, G. S. Spray-Dried Multiscale Nano-Biocomposites Containing Living Cells. *ACS Nano* **2015**, *9*, 6961–6977.

(18) Baca, H. K.; Ashley, C.; Carnes, E.; Lopez, D.; Flemming, J.; Dunphy, D.; Singh, S.; Chen, Z.; Liu, N.; Fan, H.; *et al.* Cell-Directed Assembly of Lipid-Silica Nanostructures Providing Extended Cell Viability. *Science* **2006**, *313*, 337–341.

(19) Bhatia, R. B.; Brinker, C. J.; Gupta, A. K.; Singh, A. K. Aqueous Sol-Gel Process for Protein Encapsulation. *Chem. Mater.* **2000**, *12*, 2434–2441.

(20) Naidu, C.; Suneetha, Y. Review Article: Current Knowledge on Microarray Technology - An Overview. *Trop. J. Pharm. Res.* **2012**, *11*, 153–164.

(21) Fujita, A.; Sato, J. R.; Rodrigues, L. D. O.; Ferreira, C. E.; Sogayar, M. C. Evaluating Different Methods of Microarray Data Normalization. *BMC Bioinformatics* **2006**, *7*, 469.
(22) Lim, W. K.; Wang, K.; Lefebvre, C.; Califano, A. Comparative Analysis of Microarray Normalization Procedures: Effects on Reverse Engineering Gene Networks. In *Bioinformatics*; 2007; Vol. 23.
(23) Irizarry, R. A. Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics* **2003**, *4*, 249–264.
(24) Al-Shahrour, F.; Minguez, P.; T??rraga, J.; Montaner, D.; Alloza, E.; Vaquerizas, J. M.; Conde, L.; Blaschke, C.; Vera, J.; Dopazo, J. BABELOMICS: A Systems Biology Perspective in the Functional Annotation of Genome-Scale Experiments. *Nucleic Acids Res.* **2006**, *34*.
(25) Subramanian, A.; Kuehn, H.; Gould, J.; Tamayo, P.; Mesirov, J. P. GSEA-P: A Desktop Application for Gene Set Enrichment Analysis. *Bioinformatics* **2007**, *23*, 3251–3253.
(26) Huang, D. W.; Sherman, B. T.; Lempicki, R. A. Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* **2009**, *4*, 44–57.
(27) Nishimura, D. BioCarta. *Biotech Softw. Internet Rep.* **2001**, *2*, 117–120.
(28) Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 1999, *27*, 29–34.
(29) Gene Ontology Consortium. Gene Ontology Consortium: Going Forward. *Nucleic Acids Res.* **2015**, *43*, D1049-56.
(30) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T. Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* **2000**, *25*, 25–29.
(31) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. B* **1995**, 289–300.
(32) Bland, J. M.; Altman, D. G. Multiple Significance Tests: The Bonferroni Method. *BMJ* **1995**, *310*, 170.
(33) Annis, D. H. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*; 2005; Vol. 100.
(34) Lotia, S.; Montojo, J.; Dong, Y.; Bader, G. D.; Pico, A. R. Cytoscape App Store. *Bioinformatics* **2013**, *29*, 1350–1351.
(35) Martin, A.; Ochagavia, M. E.; Rabasa, L. C.; Miranda, J.; Fernandez-de-Cossio, J.; Bringas, R. BisoGenet: A New Tool for Gene Network Building, Visualization and Analysis. *BMC Bioinformatics* **2010**, *11*, 91.
(36) Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K. P.; *et al.* STRING v10: Protein-Protein Interaction Networks, Integrated over the Tree of Life. *Nucleic Acids Res.* **2015**, *43*, D447–D452.
(37) Park, J. H.; Hong, D.; Lee, J.; Choi, I. S. Cell-in-Shell Hybrids: Chemical Nanoencapsulation of Individual Cells. *Acc. Chem. Res.* **2016**, *49*, 792–800.
(38) Orive, G.; Santos, E.; Poncelet, D.; Hernandez, R. M.; Pedraz, J. L.; Wahlberg, L. U.; Vos, P. De; Emerich, D. Cell Encapsulation: Technical and Clinical Advances. *Trends Pharmacol. Sci.* **2015**, *36*, 537–546.
(39) Ben-Yoav, H.; Melamed, S.; Freeman, A.; Shacham-Diamand, Y.; Belkin, S. Whole-Cell Biochips for Bio-Sensing: Integration of Live Cells and Inanimate Surfaces. *Crit. Rev. Biotechnol.* **2011**, *31*, 337–353.
(40) Ruiz-Hitzky, E.; Darder, M.; Aranda, P.; Ariga, K. Advances in Biomimetic and

Nanostructured Biohybrid Materials. *Adv. Mater.* **2010**, *22*, 323–336.
(41) Leonard, A.; Dandoy, P.; Danloy, E.; Leroux, G.; Meunier, C. F.; Rooke, J. C.; Su, B.-L. Whole-Cell Based Hybrid Materials for Green Energy Production, Environmental Remediation and Smart Cell-Therapy. *Chem. Soc. Rev.* **2011**, *40*, 860–885.
(42) Burgain, J.; Gaiani, C.; Linder, M.; Scher, J. Encapsulation of Probiotic Living Cells: From Laboratory Scale to Industrial Applications. *J. Food Eng.* **2011**, *104*, 467–483.
(43) Meunier, C. F.; Rooke, J. C.; Léonard, A.; Xie, H.; Su, B.-L. Living Hybrid Materials Capable of Energy Conversion and CO 2 Assimilation. *Chem. Commun.* **2010**, *46*, 3843–3859.
(44) Carballeira, J. D.; Quezada, M. A.; Hoyos, P.; Simeó, Y.; Hernaiz, M. J.; Alcantara, A. R.; Sinisterra, J. V. Microbial Cells as Catalysts for Stereoselective Red–ox Reactions. *Biotechnol. Adv.* **2009**, *27*, 686–714.
(45) Harper, J. C.; Edwards, T. L.; Savage, T.; Harbaugh, S.; Kelley-Loughnane, N.; Stone, M. O.; Brinker, C. J.; Brozik, S. M. Orthogonal Cell-Based Biosensing: Fluorescent, Electrochemical, and Colorimetric Detection with Silica-Immobilized Cellular Communities Integrated with an ITO–Glass/Plastic Laminate Cartridge. *small* **2012**, *8*, 2743–2751.
(46) Yap, F. L.; Zhang, Y. Protein and Cell Micropatterning and Its Integration with Micro/nanoparticles Assembly. *Biosens. Bioelectron.* **2007**, *22*, 775–788.
(47) Matsusaki, M.; Kadowaki, K.; Nakahara, Y.; Akashi, M. Fabrication of Cellular Multilayers with Nanometer-Sized Extracellular Matrix Films. *Angew. Chemie* **2007**, *119*, 4773–4776.
(48) Coradin, T.; Boissière, M.; Livage, J. Sol-Gel Chemistry in Medicinal Science. *Curr. Med. Chem.* **2006**, *13*, 99–108.
(49) Lee, K. Y.; Mooney, D. J. Alginate: Properties and Biomedical Applications. *Prog. Polym. Sci.* **2012**, *37*, 106–126.
(50) Drury, J. L.; Mooney, D. J. Hydrogels for Tissue Engineering: Scaffold Design Variables and Applications. *Biomaterials* **2003**, *24*, 4337–4351.
(51) Wang, S.; Guo, Z. Bio-Inspired Encapsulation and Functionalization of Living Cells with Artificial Shells. *Colloids Surfaces B Biointerfaces* **2014**, *113*, 483–500.
(52) Nassif, N.; Livage, J. From Diatoms to Silica-Based Biohybrids. *Chem. Soc. Rev.* **2011**, *40*, 849–859.
(53) Harper, J. C.; Khirpin, C. Y.; Carnes, E. C.; Ashley, C. E.; Lopez, D. M.; Savage, T.; Jones, H. D. T.; Davis, R. W.; Nunez, D. E.; Brinker, L. M. Cell-Directed Integration into Three-Dimensional Lipid− Silica Nanostructured Matrices. *ACS Nano* **2010**, *4*, 5539–5550.
(54) Conroy, J. F. T.; Power, M. E.; Martin, J.; Earp, B.; Hosticka, B.; Daitch, C. E.; Norris, P. M. Cells in Sol-Gels I: A Cytocompatible Route for the Production of Macroporous Silica Gels. *J. Sol-Gel Sci. Technol.* **2000**, *18*, 269–283.
(55) Ferrer, M. L.; Garcia-Carvajal, Z. Y.; Yuste, L.; Rojo, F.; del Monte, F. Bacteria Viability in Sol-Gel Materials Revisited: Cryo-SEM as a Suitable Tool to Study the Structural Integrity of Encapsulated Bacteria. *Chem. Mater.* **2006**, *18*, 1458–1463.
(56) Nassif, N.; Bouvet, O.; Rager, M. N.; Roux, C.; Coradin, T.; Livage, J. Living Bacteria in Silica Gels. *Nat. Mater.* **2002**, *1*, 42–44.
(57) Depagne, C.; Roux, C.; Coradin, T. How to Design Cell-Based Biosensors Using the Sol–gel Process. *Anal. Bioanal. Chem.* **2011**, *400*, 965–976.
(58) Nassif, N.; Roux, C.; Coradin, T.; Rager, M.-N.; Bouvet, O. M. M.; Livage, J. A Sol–gel Matrix to Preserve the Viability of Encapsulated Bacteria. *J. Mater. Chem.* **2003**, *13*, 203–

208.
(59) Baca, H. K.; Carnes, E.; Singh, S.; Ashley, C.; Lopez, D.; Brinker, C. J. Cell-Directed Assembly of Bio/nano Interfaces—a New Scheme for Cell Immobilization. *Acc. Chem. Res.* **2007**, *40*, 836–845.
(60) Kültz, D. Molecular and Evolutionary Basis of the Cellular Stress Response. *Annu.Rev.Physiol.* **2005**, *67*, 225–257.
(61) Luscombe, N. M.; Babu, M. M.; Yu, H.; Snyder, M.; Teichmann, S. A.; Gerstein, M. Genomic Analysis of Regulatory Network Dynamics Reveals Large Topological Changes. *Nature* **2004**, *431*, 308–312.
(62) Cvitkovitch, D. G.; Li, Y. H.; Ellen, R. P. Quorum Sensing and Biofilm Formation in Streptococcal Infections. *J. Clin. Invest.* **2003**, *112*, 1626–1632.
(63) Carturan, G.; Monte, R. D.; Pressi, G.; Secondin, S.; Verza, P. Production of Valuable Drugs from Plant Cells Immobilized by Hybrid Sol-Gel SiO2. *J. Sol-Gel Sci. Technol.* **1998**, *13*, 273–276.
(64) Eleftheriou, N. M.; Ge, X.; Kolesnik, J.; Falconer, S. B.; Harris, R. J.; Khursigara, C.; Brown, E. D.; Brennan, J. D. Entrapment of Living Bacterial Cells in Low-Concentration Silica Materials Preserves Cell Division and Promoter Regulation. *Chem. Mater.* **2013**, *25*, 4798–4805.
(65) Blondeau, M.; Brayner, R.; Guyot, F.; Coradin, T. Correlating Biological Methods to Assess Escherichia Coli Bacteria Viability in Silica Gels. *Anal. Methods* **2014**, *6*, 2429–2431.
(66) Carnes, E. C.; Lopez, D. M.; Donegan, N. P.; Cheung, A.; Gresham, H.; Timmins, G. S.; Brinker, C. J. Confinement-Induced Quorum Sensing of Individual Staphylococcus Aureus Bacteria. *Nat. Chem. Biol.* **2010**, *6*, 41–45.
(67) Nassif, N.; Roux, C.; Coradin, T.; Bouvet, O. M. M.; Livage, J. Bacteria Quorum Sensing in Silica Matrices. *J. Mater. Chem.* **2004**, *14*, 2264–2268.
(68) Harper, J. C.; Lopez, D. M.; Larkin, E. C.; Economides, M. K.; McIntyre, S. K.; Alam, T. M.; Tartis, M. S.; Werner-Washburne, M.; Brinker, C. J.; Brozik, S. M. Encapsulation of S. Cerevisiae in Poly (Glycerol) Silicate Derived Matrices: Effect of Matrix Additives and Cell Metabolic Phase on Long-Term Viability and Rate of Gene Expression. *Chem. Mater.* **2011**, *23*, 2555–2564.
(69) Depardieu, M.; Viaud, M.; Buguin, A.; Livage, J.; Sanchez, C.; Backov, R. A Multiscale Study of Bacterial Proliferation Modes within Novel E. coli@Si(HIPE) Hybrid Macrocellular Living Foams. *J. Mater. Chem. B* **2016**, *4*, 2290–2303.
(70) Heim, S.; Lleo, M.; Bonato, B.; Guzman, C. A.; Canepari, P. The Viable but Nonculturable State and Starvation Are Different Stress Responses of Enterococcus Faecalis, as Determined by Proteome Analysis. *J. Bacteriol.* **2002**, *184*, 6739–6745.
(71) Back, J. P.; Kroll, R. G. The Differential Fluorescence of Bacteria Stained with Acridine Orange and the Effects of Heat. *J. Appl. Bacteriol.* **1991**, *71*, 51–58.
(72) Davey, H. M.; Hexley, P. Red but Not Dead? Membranes of Stressed Saccharomyces Cerevisiae Are Permeable to Propidium Iodide. *Environ. Microbiol.* **2011**, *13*, 163–171.
(73) Savage, T. J.; Dunphy, D. R.; Harbaugh, S.; Kelley-Loughnane, N.; Harper, J. C.; Brinker, C. J. Influence of Silica Matrix Composition and Functional Component Additives on the Bioactivity and Viability of Encapsulated Living Cells. *ACS Biomater. Sci. Eng.* **2015**, *1*, 1231–1238.
(74) Wright, R. Transmission Electron Microscopy of Yeast. *Microsc. Res. Tech.* **2000**, *51*, 496–510.

(75) Smyth, G. K. Limma: Linear Models for Microarray Data. In; Bioinformatics and computational biology solutions using R and Bioconductor; Springer, 2005; pp. 397–420.
(76) Wu, Z.; Irizarry, R. A.; Gentleman, R.; Martinez-Murillo, F.; Spencer, F. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *J. Am. Stat. Assoc.* **2004**, *99*, 909–917.
(77) Supek, F.; Bošnjak, M.; Škunca, N.; Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* **2011**, *6*, e21800.
(78) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.
(79) Lu, Y.; Fan, H.; Stump, A.; Ward, T. L.; Rieker, T.; Brinker, C. J. Aerosol-Assisted Self-Assembly of Mesostructured Spherical Nanoparticles. *Nature* **1999**, *398*, 223–226.
(80) Brinker, C. J.; Scherer, G. W. Sol-Gel Science, 1990. *Phys. Chem. sol–gel Process.* **1990**.
(81) Wong, Y. L.; Sampson, S.; Germishuizen, W. A.; Goonesekera, S.; Caponetti, G.; Sadoff, J.; Bloom, B. R.; Edwards, D. Drying a Tuberculosis Vaccine without Freezing. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2591–2595.
(82) Brinker, C. J.; Keefer, K. D.; Schaefer, D. W.; Ashley, C. S. Sol-Gel Transition in Simple Silicates. *J. Non. Cryst. Solids* **1982**, *48*, 47–64.
(83) Brinker, C. J.; Keefer, K. D.; Schaefer, D. W.; Assink, R. A.; Kay, B. D.; Ashley, C. S. Sol-Gel Transition in Simple Silicates II. *J. Non. Cryst. Solids* **1984**, *63*, 45–59.
(84) Gray, J. V; Petsko, G. A.; Johnston, G. C.; Ringe, D.; Singer, R. A.; Werner-Washburne, M. "Sleeping Beauty": Quiescence in Saccharomyces Cerevisiae. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 187–206.
(85) Boorsma, A.; de Nobel, H.; ter Riet, B.; Bargmann, B.; Brul, S.; Hellingwerf, K. J.; Klis, F. M. Characterization of the Transcriptional Response to Cell Wall Stress in Saccharomyces Cerevisiae. *Yeast* **2004**, *21*, 413–427.
(86) Neiman, A. M. Sporulation in the Budding Yeast Saccharomyces Cerevisiae. *Genetics* **2011**, *189*, 737–765.
(87) Liu, L.; Sakakibara, K.; Chen, Q.; Okamoto, K. Receptor-Mediated Mitophagy in Yeast and Mammalian Systems. *Cell Res.* **2014**, *24*, 787–795.
(88) Manjithaya, R.; Nazarko, T. Y.; Farré, J. C.; Subramani Suresh, S. Molecular Mechanism and Physiological Role of Pexophagy. *FEBS Letters*, 2010, *584*, 1367–1373.
(89) Levin, D. E. Regulation of Cell Wall Biogenesis in Saccharomyces Cerevisiae: The Cell Wall Integrity Signaling Pathway. *Genetics* **2011**, *189*, 1145–1175.
(90) Fuchs, B. B.; Mylonakis, E. Our Paths Might Cross: The Role of the Fungal Cell Wall Integrity Pathway in Stress Response and Cross Talk with Other Stress Response Pathways. *Eukaryot. Cell* **2009**, *8*, 1616–1625.
(91) Poirier, Y.; Antonenkov, V. D.; Glumoff, T.; Hiltunen, J. K. Peroxisomal β-Oxidation—a Metabolic Pathway with Multiple Functions. *Biochim. Biophys. Acta (BBA)-Molecular Cell Res.* **2006**, *1763*, 1413–1426.
(92) Swinnen, S.; Klein, M.; Carrillo, M.; McInnes, J.; Nguyen, H. T. T.; Nevoigt, E. Re-Evaluation of Glycerol Utilization in Saccharomyces Cerevisiae: Characterization of an Isolate That Grows on Glycerol without Supporting Supplements. *Biotechnol. Biofuels* **2013**, *6*, 1.
(93) Klosinska, M. M.; Crutchfield, C. A.; Bradley, P. H.; Rabinowitz, J. D.; Broach, J. R. Yeast Cells Can Access Distinct Quiescent States. *Genes Dev.* **2011**, *25*, 336–349.
(94) LaFleur, M. D.; Kumamoto, C. A.; Lewis, K. Candida Albicans Biofilms Produce

Antifungal-Tolerant Persister Cells. *Antimicrob. Agents Chemother.* **2006**, *50*, 3839–3846.
(95) Bojsen, R.; Regenberg, B.; Gresham, D.; Folkesson, A. A Common Mechanism Involving the TORC1 Pathway Can Lead to Amphotericin B-Persistence in Biofilm and Planktonic Saccharomyces Cerevisiae Populations. *Sci. Rep.* **2016**, *6*, 21874.
(96) Bojsen, R.; Regenberg, B.; Folkesson, A. Persistence and Drug Tolerance in Pathogenic Yeast. *Curr. Genet.* **2016**, 1–4.
(97) Phillips, E.; Penate-Medina, O.; Zanzonico, P. B.; Carvajal, R. D.; Mohan, P.; Ye, Y.; Humm, J.; Gönen, M.; Kalaigian, H.; Schöder, H.; *et al.* Clinical Translation of an Ultrasmall Inorganic Optical-PET Imaging Nanoparticle Probe. *Sci. Transl. Med.* **2014**, *6*, 260ra149.
(98) Zhang, Y. Persisters, Persistent Infections and the Yin–Yang Model. *Emerg. Microbes Infect.* **2014**, *3*, e3.
(99) Maisonneuve, E.; Gerdes, K. Molecular Mechanisms Underlying Bacterial Persisters. *Cell*, 2014, *157*, 539–548.
(100) Fazal, Z.; Pelowitz, J.; Johnson, P. E.; Harper, J. C.; Brinker, C. J.; Jakobsson, E. Three-Dimensional Encapsulation of Saccharomyces Cerevisiae in Silicate Matrices Creates Distinct Metabolic States as Revealed by Gene Chip Analysis. *ACS Nano 0*, null.
(101) Yamada, T.; Letunic, I.; Okuda, S.; Kanehisa, M.; Bork, P. IPath2.0: Interactive Pathway Explorer. *Nucleic Acids Res.* **2011**, *39*.
(102) Acker, H. Van; Dijck, P. Van; Coenye, T. Molecular Mechanisms of Antimicrobial Tolerance and Resistance in Bacterial and Fungal Biofilms. *Trends Microbiol.* **2014**, *22*, 326–333.
(103) Cyert, M. S.; Philpott, C. C. Regulation of Cation Balance in Saccharomyces Cerevisiae. *Genetics* **2013**, *193*, 677–713.
(104) Lesuisse, E.; Raguzzi, F.; Crichton, R. R. Iron Uptake by the Yeast Saccharomyces Cerevisiae: Involvement of a Reduction Step. *Microbiology* **1987**, *133*, 3229–3236.
(105) Klionsky, D. J.; Herman, P. K.; Emr, S. D. The Fungal Vacuole: Composition, Function, and Biogenesis. *Microbiol. Rev.* **1990**, *54*, 266–292.
(106) Keren, I.; Shah, D.; Spoering, A.; Kaldalu, N.; Lewis, K. Specialized Persister Cells and the Mechanism of Multidrug Tolerance in Escherichia Coli. *J. Bacteriol.* **2004**, *186*, 8172–8180.
(107) Rodrigues, C. F.; Rodrigues, M. E.; Silva, S.; Henriques, M. Candida Glabrata Biofilms: How Far Have We Come? *J. Fungi* **2017**, *3*.
(108) Huang, D. W.; Sherman, B. T.; Lempicki, R. A.; Huang, D.; Sherman, B.; Lempicki, R.; Huang, D.; Sherman, B.; Lempicki, R.; Maere, S.; *et al.* Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic Acids Res.* **2009**, *37*, 1–13.
(109) Eden, E.; Navon, R.; Steinfeld, I.; Lipson, D.; Yakhini, Z. GOrilla: A Tool for Discovery and Visualization of Enriched GO Terms in Ranked Gene Lists. *BMC Bioinformatics* **2009**, *10*, 48.
(110) Reimand, J.; Arak, T.; Adler, P.; Kolberg, L.; Reisberg, S.; Peterson, H.; Vilo, J. g:Profiler—a Web Server for Functional Interpretation of Gene Lists (2016 Update). *Nucleic Acids Res.* **2016**, *44*, W83–W89.
(111) Conesa, A.; Götz, S.; García-Gómez, J. M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A Universal Annotation and Visualization Tool in Functional Genomics Research. Application Note. *Bioinformatics* **2005**, *21*, 3674–3676.
(112) Zeeberg, B. R.; Qin, H.; Narasimhan, S.; Sunshine, M.; Cao, H.; Kane, D. W.; Reimers, M.;

Stephens, R. M.; Bryant, D.; Burt, S. K.; *et al.* High-Throughput GoMiner, an "Industrial-Strength" Integrative Gene Ontology Tool for Interpretation of Multiple-Microarray Experiments, with Application to Studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics* **2005**, *6*, 168.

(113) Al-Shahrour, F.; Díaz-Uriarte, R.; Dopazo, J. FatiGO: A Web Tool for Finding Significant Associations of Gene Ontology Terms with Groups of Genes. *Bioinformatics* **2004**, *20*.

(114) Ballouz, S.; Pavlidis, P.; Gillis, J. Using Predictive Specificity to Determine When Gene Set Analysis Is Biologically Meaningful. *Nucleic Acids Res.* **2016**, gkw957.

(115) Alexa, A.; Rahnenführer, J.; Lengauer, T. Improved Scoring of Functional Groups from Gene Expression Data by Decorrelating GO Graph Structure. *Bioinformatics* **2006**, *22*, 1600–1607.

(116) Alexa, A.; Rahnenfuhrer, J. topGO: topGO: Enrichment Analysis for Gene Ontology. R Package Version 2.18.0. *October* **2010**.

(117) Falcon, S.; Gentleman, R. Using GOstats to Test Gene Lists for GO Term Association. *Bioinformatics* **2007**, *23*, 257–258.

(118) Maere, S.; Heymans, K.; Kuiper, M. BiNGO: A Cytoscape Plugin to Assess Overrepresentation of Gene Ontology Categories in Biological Networks. *Bioinformatics* **2005**, *21*, 3448–3449.

(119) Rivals, I.; Personnaz, L.; Taing, L.; Potier, M.-C. Enrichment or Depletion of a GO Category within a Class of Genes: Which Test? *Bioinformatics* **2007**, *23*, 401.

(120) Benjamini, Y.; Yekutieli, D. The Control of the False Discovery Rate in Multiple Testing under Depencency. *Ann. Stat.* **2001**, *29*, 1165–1188.

(121) Zheng, Q.; Wang, X. J. GOEAST: A Web-Based Software Toolkit for Gene Ontology Enrichment Analysis. *Nucleic Acids Res.* **2008**, *36*.

(122) Blüthgen, N.; Brand, K.; Cajavec, B.; Swat, M.; Herzel, H.; Beule, D. Biological Profiling of Gene Groups Utilizing Gene Ontology. *Genome Informatics* **2005**, *16*, 106–115.

(123) POWERS, D. M. W. Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.

(124) Yon Rhee, S.; Wood, V.; Dolinski, K.; Draghici, S. Use and Misuse of the Gene Ontology Annotations. *Nat. Rev. Genet.* **2008**, *9*, 509–515.

(125) Alaux, C.; Sinha, S.; Hasadsri, L.; Hunt, G. J.; Guzman-Novoa, E.; DeGrandi-Hoffman, G.; Uribe-Rubio, J. L.; Southey, B. R.; Rodriguez-Zas, S.; Robinson, G. E. Honey Bee Aggression Supports a Link between Gene Regulation and Behavioral Evolution. *Proc. Natl. Acad. Sci.* **2009**, *106*, 15400–15405.

(126) Liu, H.; Robinson, G. E.; Jakobsson, E. Conservation in Mammals of Genes Associated with Aggression-Related Behavioral Phenotypes in Honey Bees. *PLoS Comput. Biol.* **2016**, *12*.

Appendix A: GO terms and KEGG pathways enriched in each encapsulation method

Table S1 Complete list of GO terms enriched in CDA at FDR-adjusted P-value <0.05
Table S2 Complete list of GO terms enriched in AqS+g at FDR-adjusted P-value <0.05
Table S3 Complete list of GO terms enriched in SD at FDR-adjusted P-value <0.05
Table S4 Complete list of KEGG pathways enriched in CDA, AqS+g and SD at FDR-adjusted P-value <0.05

Appendix B: Workflow for TopGO and GOstats automated pipeline

A TopGO- and GOstats-based automated pipeline for GO enrichment analysis using F-measure optimization based on resampling and traditional calculation.