UNIVERSITY OF CALIFORNIA, SAN DIEGO

Development of experimental and computational tools for high-throughput microbial "omics"

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Gordon Jeff Bean

Committee in charge:

Professor Trey Ideker, Chair
Professor Nuno Bandeira
Professor Charles Elkan
Professor Bruce Hamilton
Professor Dong Wang

2014

The Dissertation of Gordon Jeff Bean is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2014

DEDICATION

To God, my loving wife, and all those who believed I could make it.

EPIGRAPH

Keep Trying. Be believing. Be happy. Don't get discouraged. Things will work out.

- Gordon B. Hinckley

Good judgment comes from experience. Experience comes from bad judgment.

- Jeff Bean

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

DDR	DNA damage response
EMAP	Epistasis mini-array profiling / Epistasis mapping
FDR	False discovery rate
GO	Gene Ontology
MMS	Methylmethane sulfonate
NER	Nucleotide excision repair
PPI	Protein-protein interaction
SC	Synthetic complete
SGA	Synthetic genetic array
UV	Ultra-violate
Y2H	Yeast 2-hybrid
YPAD	Yeast peptone adenine dextrose

LIST OF FIGURES AND TABLES

ACKNOWLEDGMENTS

I would like to acknowledge Professor Trey Ideker for his invaluable guidance, encouragement, and patience as my advisor and committee chair.

I would also like to acknowledge the fantastic members of the Ideker Lab. They are good friends and helpful colleagues. I especially acknowledge Philipp Jaeger for his kindness and encouragement, as well as his scientific advice and collaboration.

Chapter 1, in full, is a reprint of the material as it appears in Differential analysis of high-throughput quantitative genetic interaction data. Bean, G. J. & Ideker, T. Genome Biology. 13, R123 (2012). The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Development of ultra-high-density screening tools for microbial "omics". Bean, G. J., Jaeger, P. a, Bahr, S. & Ideker, T. PLoS One 9, e85177 (2014). The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Bean, Gordon; Jaeger, Philipp; Ideker, Trey. The dissertation author was the primary investigator and author of this material.

VITA

2008 Bachelor of Science, Brigham Young University

2014 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Bean, G. J. & Ideker, T. Differential analysis of high-throughput quantitative genetic interaction data. *Genome Biol.* **13,** R123 (2012).

Bean, G.J., Dimarco, E.A., Mercer, L.D., Thayer, L.K., Roy, A., & Ghosal, S. Finite skew-mixture models for estimation of positive false discovery rates. *Stat. Methodol.* **10,** 46–57 (2013).

Bean, G. J., Jaeger, P. a, Bahr, S. & Ideker, T. Development of ultra-high-density screening tools for microbial "omics". *PLoS One* **9,** e85177 (2014).

ABSTRACT OF THE DISSERTATION

Development of experimental and computational tools for high-throughput microbial "omics"

by

Gordon Jeff Bean

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California, San Diego, 2014

Professor Trey Ideker, Chair

As the study of biology has increasingly utilized high-throughput experimental platforms, the need for rigorous, computational analysis of systems biology data has never been greater. Here I present three aims

towards solving important computational problems in the field of high-throughput genetic interaction mapping and screening.

In Chapter 1, I provide a reproduction of published work in which we present a strategy that leverages statistical information from the experimental design to produce a novel, quantitative differential interaction score, which performs favorably compared to previous differential scores. We also discuss the added utility of differential genetic-similarity in differential network analysis.

In Chapter 2, I present previously published work in which we present an ultra-high-density, 6144-colony arraying system and analysis toolbox. Using budding yeast as a benchmark, we find that these tools boost genetic screening throughput 4-fold and yield significant cost and time reductions at quality levels equal to or better than current methods. We conclude that the new ultra-high-density screening tools enable researchers to significantly increase the size and scope of their genetic screens.

In Chapter 3, I present work being prepared for publication, in which we utilize the new 6144-colony agar format to perform genome-wide time-lapse analysis of the yeast gene deletion collection to quantify dynamic growth phenotypes and identify key biological processes involved in adaptation to metabolic stress through nutrient depletion. We

then apply our method to study the dynamic response to UV-radiation and observe that growth profiles recapitulate key biological processes in DNA repair and suggest novel relationships.

As a whole, this thesis is a combined work of both computational and biological research. In Aim 1, I develop a new statistical score and make biological insights from genetic interaction profiles. In Aim 2, I developed an image analysis toolkit that supports the emerging 6144-colony format. In Aim 3, I use the new 6144 format and image analysis toolkit to develop a novel experimental platform using time-lapse imaging of yeast on agar. We use this platform to identify key cellular pathways involved in adaptation to standard growth conditions and suggest important relationships in the UV-radiation response.

CHAPTER 1

INTRODUCTION

Systems biology is founded on the belief that the miracle of life can be modeled and understood as a composite of many interacting systems – networks of cellular components and their relationships. Key to creating these models is the knowledge of the components – the genes, proteins, metabolites, etc. – and the interactions that relate them to one another and describe their activity. In its quest to answer these questions, systems biology is characterized by the use of high-throughput technologies, making dozens to thousands to millions of measurements in single experimental efforts. In consequence of the overwhelming amount of information being generated in pursuit of biological understanding, the need for bioinformatics solutions has never been greater.

Saccharomyces cerevisiae, or "baker's yeast," has been an effective model organism for studying eukaryotic cellular systems. Molecular methods for genetic manipulation of yeast are well established and inexpensive, making yeast an ideal candidate for research. Additionally, from early on the relevance of understanding biological systems in the context of yeast has been recognized, and what is learned

from this model organism provides important information about other species [1,2].

In this thesis, I describe the systems biology tools and informatics solutions I have developed in the field of genetic screening and interaction mapping in the model organism *Saccharomyces cerevisiae*. In this chapter I introduce the context of my work and explain the technologies upon which my work was built.

1.1 The yeast gene deletion collection

In the past 20 years, many genomic technologies have been developed for performing systems biology experiments in Saccharomyces cerevisiae [3]. One of these tools is the yeast gene deletion collection [4,5]. Through the use of homologous recombination, the KanMX4 drug-resistance marker, flanked by barcoding sequences, was systematically inserted into most known open reading frames (ORFs), creating haploid deletion strains for 4,757 non-essential genes (Figure 1.1). More recently, a "decreased abundance by mRNA perturbation" (DAmP) collection for essential genes was created, in which the insertion of the KanMX4 cassette at the 3' end of the gene locus disrupts transcription and causes a non-lethal down-regulation of the gene product [6]. These collections of mutant strains have been an invaluable resource in systems biology.

1.2 Genetic interactions

Genetic interactions have been defined in many ways, and different experimental platforms are each uniquely adapted to quantify or observe genetic interactions according to different criteria. In general, a genetic interaction is observed when the double mutant phenotype differs from what is expected given the two single mutant phenotypes alone [7]. In the field of high-throughput genetic interaction mapping in yeast, the primary observed phenotype is relative growth rate (growth rate of the mutant relative to growth rate of the control, often summarized as colony size or "fitness"), and the expected double mutant phenotype is the product of the two single mutant phenotypes [8,9]. The specific relationships between the single and double mutant phenotypes can sometimes provide a more nuanced interpretation of the interaction [10–12] (Figure 1.2), but in most cases simply observing whether the double mutant grows faster or slower than expected is all that is reported [13].

1.3 High-throughput genetic interaction mapping in yeast

While the investigation of genetic interactions in yeast double mutants had been performed for many years previous, in 2001 Tong et al. developed a series of pinning procedures to efficiently generate double mutants strains using the gene deletion collection, dramatically increasing the throughput of genetic interaction screens [14]. Furthermore, by using

an agar format, instead of liquid culture, colony fitness could be quantified using inexpensive imaging equipment, as compared to more involved methods such as microarray hybridization [15,16].

In brief, in the procedure described by Tong et al., a haploid query strain is created containing a mutation (i.e. gene deletion) of the MAT**α** mating type. This strain is crossed with a collection of other haploid mutants of the MAT**a** mating type, known as the "array". The resulting diploids are sporulated and then passaged through as series of selection steps, ending with the selection of haploid double mutants. This final selection is accomplished because both the query and the array have respective drug-resistance markers inserted at the locus of the deleted gene. Only spores inheriting the deletion allele from the query and the array will have the two drug-resistance markers. This high-throughput mating/selection process was termed "synthetic genetic array" (SGA; Figure 1.3). Tong et al. applied the SGA approach to screen for synthetic lethal pairs – double mutants of which the respective single mutants are viable, the but the combination is not [14,17] (see also [13,18] for reviews on early high-throughput genetic interaction screening).

In 2005, Schuldiner et al. introduced two important developments to the SGA paradigm [6]. First, they created special hypomorphic alleles of essential genes, called "decreased abundance by mRNA perturbation"

(DAmP) alleles. In these strains, the drug-resistance marker and associated promoter are inserted just upstream of the gene in question. The placement of the resistance cassette into the promoter region of the gene decreases mRNA stability, effectively reducing the quantity of the gene in the cell. They demonstrated that when the deletion allele is inviable, these special alleles are effective surrogates for deletion alleles for probing gene functions.

The second important contribution, described in greater detail a year later in Collins et al. 2006, was a quantitative scoring pipeline that accurately measured expected double mutant growth rates in order to identify double mutants that exhibited epistatic, interactions [19] (see also [20]). This important computational development greatly expanded the potential of the SGA approach, allowing it to accurately identify both positive (suppressing, rescue, and masking) as well as negative (synthetic sick and synthetic lethal) interactions. This quantitative approach was termed "epistatic mini-array profiling" (EMAP). The authors of the EMAP approach used this system to study relatively small gene collections (hundreds of genes) belonging to specific pathways or cellular components. Since the initial study of the early secretory pathway [6], the EMAP approach has been used to study chromatin remodeling [21],

kinase/phosphatase signaling [22], and transcription factor organization [23], among others.

The discovery of complex networks of genetic interactions begged the availability of a genome-wide genetic interaction map, informing not only how members of specific pathways interact, but also showing the relationship across the cellular landscape. In 2010, Costanzo et al. presented such a resource, crossing approximately 1,000 queries with over 4,000 arrays [24]. They found that genetic interaction profiles – i.e. the vector of interaction scores for all tested gene pairs of a given gene – clustered according to functional groups. Using some of the key ideas in the EMAP computational pipeline, these authors presented additional computational improvements to SGA analysis that allowed them to successful process the large dataset they created [25].

Throughout the development of the genetic interaction mapping field, computational tools were designed to analyze the specific images and data of each respective group. Because each tool was designed with a specific workflow and experimental format in mind, subsequent studies in other labs – with different needs and setups – were largely required to develop computational tools de novo. This has lead to a great abundance of yeast-on-agar image analysis and statistical analysis tools, each specializing in some particular format or need[26–31].

1.4 Differential epistasis mapping

As the SGA platform matured, with technological improvements increasing ease of use and throughput, SGA studies transitioned from mapping interactions in standard conditions to identifying genetic network changes in novel growth environments, as had been done in lower-throughput platforms before [11]. Just weeks before the landmark study of Costanzo et al., Bandyopadhyay et al. presented an EMAP study comparing the topology of genetic networks in two growth conditions – with and without the addition of methyl methanesulfonate (MMS), a DNA-damaging agent [32]. They discovered that many genetic interactions appeared, disappeared, or even changed sign when the yeast were grown under DNA-damage stress. This differential epistasis-mapping (d-EMAP) approach opened the door to many new studies probing the dynamics of genetic relationships. Several other studies have investigated genetic network modifications in the presence of other DNA-damaging stresses [33,34]. Other studies are underway using the dEMAP approach to study autophagy and telomere-deficient aging.

The development of the dEMAP pipeline posed a new computational challenge. The EMAP interaction-scoring paradigm had been developed with single-condition, unpaired data in mind. The early dEMAP studies relied upon an ad-hoc approach to quantify interesting

changes in the genetic networks. In Chapter 2, I present published work in which we develop a statistically rigorous, differential interaction score. We also demonstrate how this new score can be further analyzed to reveal additional insights to stress-response genetic effects.

1.5 Technological developments in SGA/EMAP technology

As new technologies became available, the quality and throughput of genetic interaction mapping has improved. Early studies used the 384-colony format, similar to other platforms such as liquid and liquid-to-agar. However, researchers quickly discovered that one of the primary advantages of the agar-to-agar platform was its scalability, and studies began using 768- and 1536-colony formats. As these new formats were used, computational tools had to be adapted to handle the image analysis and data processing.

In Chapter 3, I present published work in which we assess the quality and performance of the new 6144-colony format and provide the computational tools needed for image analysis and data processing of the new format.

1.6 Growth-curve analysis

Growth curve analysis has been used for many years in liquid and liquid-to-agar platforms in order to determine or quantify the growth rate of mutant strains (reviewed in [35]). To summarize, in these studies, multiple

measurements of colony density, quantified either through liquid optical density (OD) or sum colony pixel intensity (SCPI) readouts. Many approaches exist for extracting the growth-rate information from growth curve data. Some fit standard-form logistic functions to the measured data, and the growth rate is directly determined from the parameters of the fitted function, whereas others use custom, ad-hoc approaches that are better adapted to the specific curve shapes and context of the experiment at hand.

Because both liquid-to-agar and agar-to-agar platforms use image analysis of yeast on agar to quantify growth, there are important similarities between the two platforms. In both cases, colony recognition/labeling and pixel thresholding algorithms are needed, and various methods exist for each unique experimental approach [30,28]. However, important differences distinguish liquid-to-agar from agar-to-agar experiments. In liquid-to-agar, colonies are pinned using broad, flat pins dipped in liquid culture. This leaves an invisible circular region of inoculated agar. Over time, the opacity of this region increases as the yeast grow, while the overall area does not increase as much. Thus, measuring colony opacity is key to tracking growth. This differs from agar-to-agar platforms. Here, yeast material is transferred from one agar plat to another, inoculating the fresh plate with semi-formed, visible colonies.

While initially a little transparent, these colonies achieve full opacity within hours of pinning (compared to days in the liquid-to-agar case) and the growth is manifested in the increase of colony area. Because growth in the liquid-to-agar setting can occur in 2 dimensions (increasing area and increasing depth), the growth dynamics are super-linear. However, in agar-to-agar platforms, the growth rate is directly proportional to the radius of the colony.

While in most studies based on growth curve analysis, only the growth rate is measured, some studies have shown that other physiological phenotypes can be easily measured from growth curve data, such as length of lag phase or total growth efficiency, and that these phenotypes convey important information growth rate alone does not [36,37]. These studies, however, have been limited to the 96-well format, using the liquid OD platform, and are limited in coverage of both mutants and conditions.

There is good precedent for performing growth curve analysis in genetic screens, and there is strong evidence that useful information besides growth rate can be extracted, but to date no one as leveraged the super-high-throughput capacity of agar-to-agar platforms to perform genome-wide growth characterizations. In Chapter 4, I present a novel experimental platform for quantifying growth curve information using the

new 6144-colony format and time-lapse imaging. This new platform allows for screening the entire yeast gene deletion collection and DAmP collection in a single agar plate, which permits performing genome-wide screens across a multitude of genetic backgrounds or growth media without increasing experimental costs or overhead.

1.7 Figures

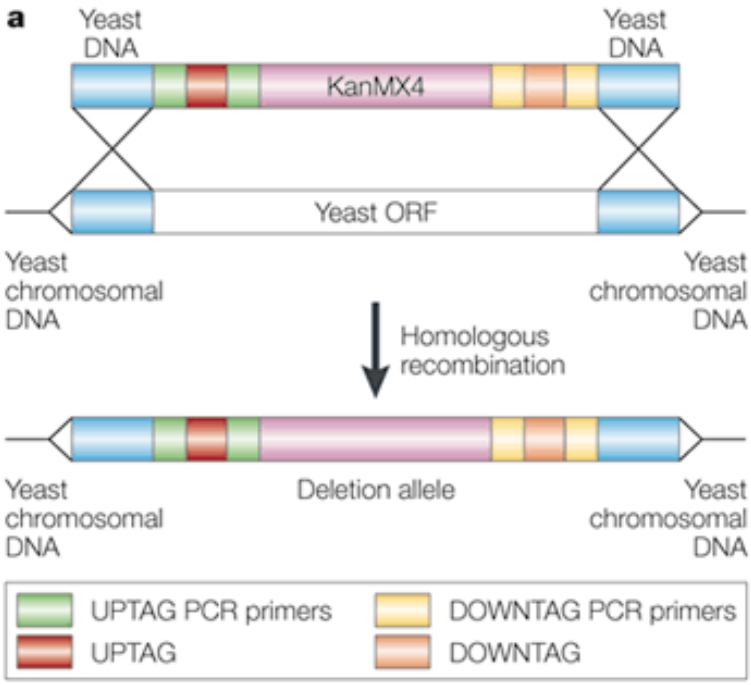
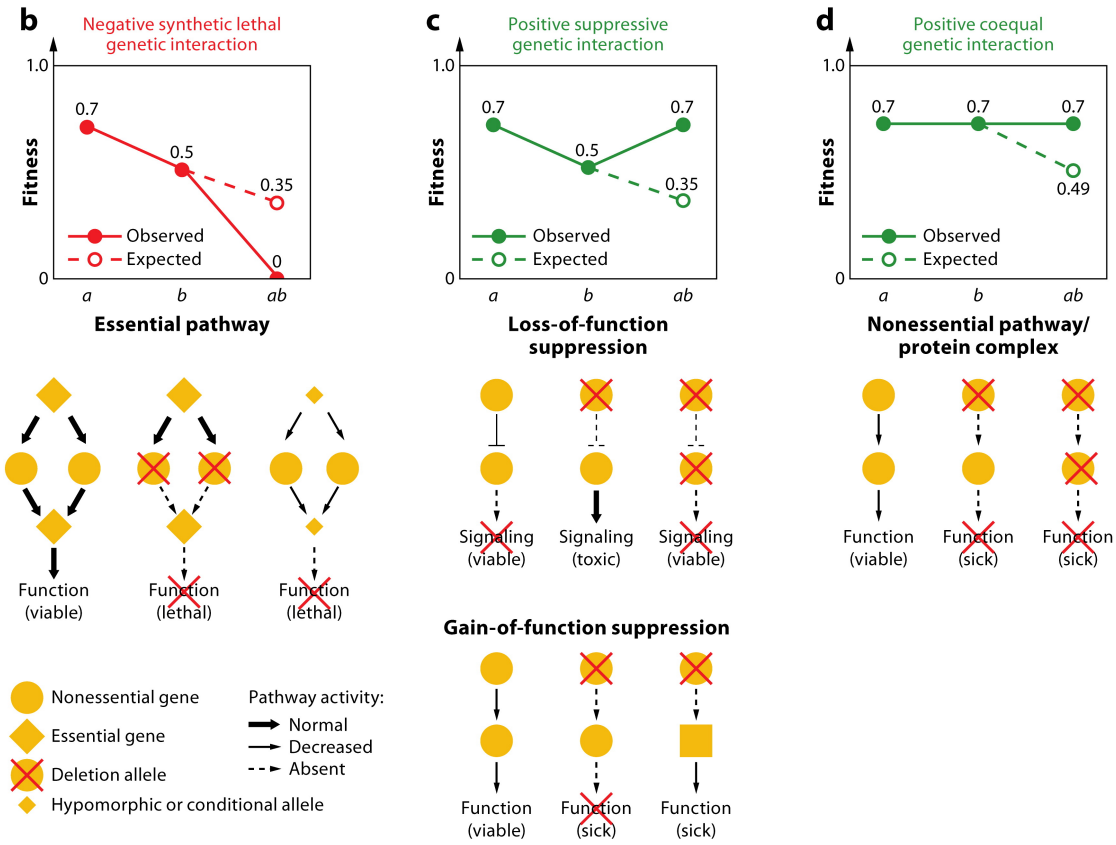


Figure 1.1 Development of the yeast gene deletion collection.
Deletion cassettes are generated by PCR such that each cassette is flanked by two 45-bp regions of yeast DNA sequence that correspond to the intended deletion target. These short regions of homology direct the integration of the deletion cassette to its intended genomic locus, resulting in a precise start-to-stop codon gene replacement. (ORF, open reading frame); KanMX4, marker that confers resistance to the antibiotic geneticin (G418).) Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics. "Emerging technologies in yeast genetics." Nature Reviews Genetics 2, 302-312 (April 2001), copyright 2001{Kumar 2001}.



Baryshnikova A, et al. 2013.
Annu. Rev. Genomics Hum. Genet. 14:111–33

Figure 1.2 Classic interpretations of negative and positive genetic interactions.
(b) Example of a negative synthetic lethal genetic interaction, whereby the observed fitness of the double mutant (*ab* = 0.0) is lower than expected (*ab* = 0.35). **(c)** Example of a positive suppressive genetic interaction, whereby the observed fitness of the double mutant (*ab* = 0.7) is greater than expected (*ab* = 0.35) and greater than the sickest of the single mutants (*b* = 0.5). **(d)** Example of a positive coequal genetic interaction, whereby the fitness phenotypes of the two single mutants and the double mutant are quantitatively indistinguishable (*a* = *b* = *ab* = 0.7). Figure and legend adapted from Baryshnikova, A., Costanzo, M., Myers, C. L., Andrews, B. & Boone, C. Genetic interaction networks: toward an understanding of heritability. Annu. Rev. Genomics Hum. Genet. 14, 111–33 (2013).

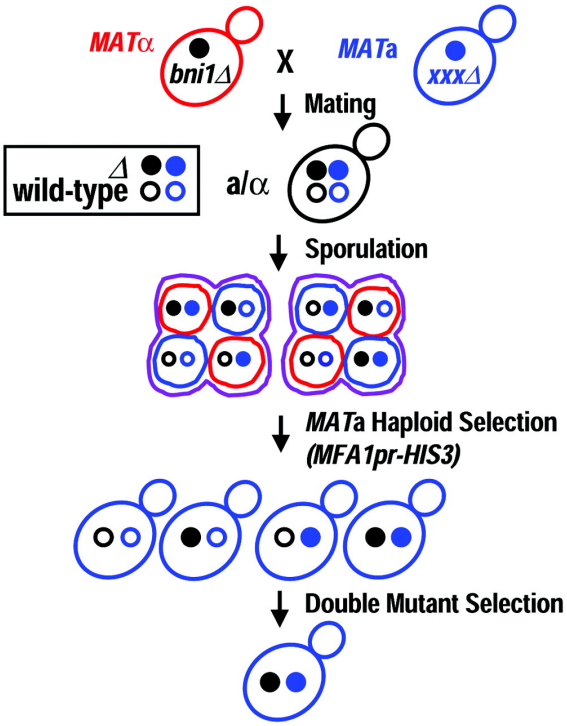


Figure 1.3 Synthetic genetic array methodology.
(i) A MATα strain carrying a query mutation (bni1Δ) linked to a dominant selectable marker, such as the nourseothricin-resistance marker natMX that confers resistance to the antibiotic nourseothricin, and anMFA1pr-HIS3 reporter is crossed to an ordered array ofMAT a viable yeast deletion mutants, each carrying a gene deletion mutation linked to a kanamycin-resistance marker (kanMX). Growth of resultant heterozygous diploids is selected for on medium containing nourseothricin and kanamycin. (ii) The heterozygous diploids are transferred to medium with reduced levels of carbon and nitrogen to induce sporulation and the formation of haploid meiotic spore progeny. (iii) Spores are transferred to synthetic medium lacking histidine, which allows for selective germination of MAT a meiotic progeny because these cells express the MFA1pr-HIS3 reporter specifically. (iv) The MAT a meiotic progeny are transferred to medium that contains both nourseothricin and kanamycin, which then selects for growth of double-mutant meiotic progeny. From Tong, a H. et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294, 2364–8 (2001). Reprinted with permission from AAAS.

1.8 References

1. Botstein D, Chervitz S a, Cherry JM: Yeast as a model organism. *Science* 1997, 277:1259–60.

2. Forsburg SL: The art and design of genetic screens: yeast. *Nat. Rev. Genet.* 2001, 2:659–6810.1038/35088500.

3. Kumar a, Snyder M: Emerging technologies in yeast genomics. *Nat. Rev. Genet.* 2001, 2:302–1210.1038/35066084.

4. Winzeler EA1, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Véronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW: Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* 1999, 285:901–610.1126/science.285.5429.901.

5. Scherens B, Goffeau A: The uses of genome-wide yeast mutant collections. *Genome Biol.* 2004, 5:22910.1186/gb-2004-5-7-229.

6. Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, Weissman JS, Krogan NJ: Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 2005, 123:507–1910.1016/j.cell.2005.08.031.

7. Phillips PC: Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 2008, 9:855–6710.1038/nrg2452.

8. Segrè D, Deluna A, Church GM, Kishony R: Modular epistasis in yeast metabolism. *Nat. Genet.* 2005, 37:77–8310.1038/ng1489.

9. Mani R, St Onge RP, Hartman JL, Giaever G, Roth FP: Defining genetic interaction. *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105:3461–610.1073/pnas.0712255105.

10. Drees BL, Thorsson V, Carter GW, Rives AW, Raymond MZ, Avila-Campillo I, Shannon P, Galitski T: Derivation of genetic interaction networks from quantitative phenotype data. *Genome Biol.* 2005, 6:R3810.1186/gb-2005-6-4-r38.

11. St Onge RP, Mani R, Oh J, Proctor M, Fung E, Davis RW, Nislow C, Roth FP, Giaever G: Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* 2007, 39:199–20610.1038/ng1948.

12. Baryshnikova A, Costanzo M, Myers CL, Andrews B, Boone C: Genetic interaction networks: toward an understanding of heritability. *Annu. Rev. Genomics Hum. Genet.* 2013, 14:111–3310.1146/annurev-genom-082509-141730.

13. Boone C, Bussey H, Andrews BJ: Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* 2007, 8:437–4910.1038/nrg2085.

14. Tong a H, Evangelista M, Parsons a B, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C: Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 2001, 294:2364–810.1126/science.1065810.

15. Pan X, Yuan DS, Ooi S-L, Wang X, Sookhai-Mahadeo S, Meluh P, Boeke JD: dSLAM analysis of genome-wide genetic interactions in Saccharomyces cerevisiae. *Methods* 2007, 41:206–2110.1016/j.ymeth.2006.07.033.

16. Ooi SL, Shoemaker DD, Boeke JD: DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. *Nat. Genet.* 2003, 35:277–8610.1038/ng1258.

17. Tong AH1, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Ménard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C: Global mapping of the yeast genetic interaction network. *Science* 2004, 303:808–1310.1126/science.1091317.

18. Ooi SL, Pan X, Peyser BD, Ye P, Meluh PB, Yuan DS, Irizarry R a, Bader JS, Spencer F a, Boeke JD: Global synthetic-lethality analysis and yeast functional profiling. *Trends Genet.* 2006, 22:56–6310.1016/j.tig.2005.11.003.

19. Collins SR, Schuldiner M, Krogan NJ, Weissman JS: A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* 2006, 7:R6310.1186/gb-2006-7-7-r63.

20. Collins SR, Roguev A, Krogan NJ: Quantitative Genetic Interaction Mapping Using the E-MAP Approach. *Methods Enzymol.* 2010, 687910.1016/S0076-6879(10)70009-4.Quantitative.

21. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, Schuldiner M, Gebbia M, Recht J, Shales M, Ding H, Xu H, Han J, Ingvarsdottir K, Cheng B, Andrews B, Boone C, Berger SL, Hieter P, Zhang Z, Brown GW, Ingles CJ, Emili A, Allis CD, Toczyski DP, Weissman JS, Greenblatt JF, Krogan NJ: Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 2007, 446:806–1010.1038/nature05649.

22. Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, Silva AC, Shales M, Collins SR, van Wageningen S, Kemmeren P, Holstege FCP, Weissman JS, Keogh M-C, Koller D, Shokat KM, Krogan NJ: Functional organization of the S. cerevisiae phosphorylation network. *Cell* 2009, 136:952–6310.1016/j.cell.2008.12.039.

23. Zheng J, Benschop JJ, Shales M, Kemmeren P, Greenblatt J, Cagney G, Holstege

F, Li H, Krogan NJ: Epistatic relationships reveal the functional organization of yeast transcription factors. *Mol. Syst. Biol.* 2010, 6:42010.1038/msb.2010.77.

24. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Tong AH, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C.: The genetic landscape of a cell. *Science* 2010, 327:425–3110.1126/science.1180823.

25. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, Youn J-Y, Ou J, San Luis B-J, Bandyopadhyay S, Hibbs M, Hess D, Gingras A-C, Bader GD, Troyanskaya OG, Brown GW, Andrews B, Boone C, Myers CL: Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* 2010, 7:1017–2410.1038/nmeth.1534.

26. Dittmar JC, Reid RJ, Rothstein R: ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data. *BMC Bioinformatics* 2010, 11:35310.1186/1471-2105-11-353.

27. Wagih O, Usaj M, Baryshnikova A, Vandersluis B, Kuzmin E, Costanzo M, Myers CL, Andrews BJ, Boone CM, Parts L: SGAtools: one-stop analysis and visualization of array-based genetic interaction screens. *Nucleic Acids Res.* 2013, :1–610.1093/nar/gkt400.

28. Shah N a, Laws RJ, Wardman B, Zhao LP, Hartman JL: Accurate, precise modeling of cell proliferation kinetics from time-lapse imaging and automated image analysis of agar yeast culture arrays. *BMC Syst. Biol.* 2007, 1:310.1186/1752-0509-1-3.

29. Memarian N, Jessulat M, Alirezaie J, Mir-Rashed N, Xu J, Zareie M, Smith M, Golshani A: Colony size measurement of the yeast gene deletion strains for functional genomics. *BMC Bioinformatics* 2007, 8:11710.1186/1471-2105-8-117.

30. Lawless C, Wilkinson DJ, Young A, Addinall SG, Lydall D a: Colonyzer: automated quantification of micro-organism growth characteristics on solid agar. *BMC Bioinformatics* 2010, 11:28710.1186/1471-2105-11-287.

31. Hartman JL, Tippery NP: Systematic quantification of gene interactions by phenotypic array analysis. *Genome Biol.* 2004, 5:R4910.1186/gb-2004-5-7-r49.

32. Bandyopadhyay S, Mehta M, Kuo D, Sung M-K, Chuang R, Jaehnig EJ, Bodenmiller B, Licon K, Copeland W, Shales M, Fiedler D, Dutkowski J, Guénolé A, van Attikum H, Shokat KM, Kolodner RD, Huh W-K, Aebersold R, Keogh M-C, Krogan NJ, Ideker T: Rewiring of genetic networks in response to DNA damage. *Science* 2010, 330:1385–910.1126/science.1195618.

33. Guénolé A, Srivas R, Vreeken K, Wang ZZ, Wang S, Krogan NJ, Ideker T, van Attikum H: Dissection of DNA damage responses using multiconditional genetic interaction

maps. *Mol. Cell* 2013, 49:346–5810.1016/j.molcel.2012.11.023.

34. Srivas R, Costelloe T, Carvunis A-R, Sarkar S, Malta E, Sun SM, Pool M, Licon K, van Welsem T, van Leeuwen F, McHugh PJ, van Attikum H, Ideker T: A UV-induced genetic network links the RSC complex to nucleotide excision repair and shows dose-dependent rewiring. *Cell Rep.* 2013, 5:1714–2410.1016/j.celrep.2013.11.035.

35. Blomberg A: Measuring growth rate in high-throughput growth phenotyping. *Curr. Opin. Biotechnol.* 2011, 22:94–10210.1016/j.copbio.2010.10.013.

36. Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A: High-resolution yeast phenomics resolves different physiological features in the saline response. *Proc. Natl. Acad. Sci. U. S. A.* 2003, 100:15724–910.1073/pnas.2435976100.

37. Warringer J, Anevski D, Liu B, Blomberg A: Chemogenetic fingerprinting by analysis of cellular growth dynamics. *BMC Chem. Biol.* 2008, 8:310.1186/1472-6769-8-3.

CHAPTER 2

DIFFERENTIAL ANALYSIS OF HIGH-THROUGHPUT QUANTITATIVE GENETIC INTERACTION DATA

Gordon Bean and Trey Ideker

2.1 Abstract

Synthetic genetic arrays (SGA) have been very effective at measuring genetic interactions in yeast in a high throughput manner and recently have been expanded to measure quantitative changes in interaction, termed 'differential interactions', across multiple conditions. Here, we present a strategy that leverages statistical information from the experimental design to produce a novel, quantitative differential interaction score, which performs favorably compared to previous differential scores. We also discuss the added utility of differential genetic-similarity in differential network analysis. Our approach is preferred for differential network analysis, and our implementation, written in MATLAB, can be found at http://chianti.ucsd.edu/~gbean/compute_differential_scores.m.

2.2 Background

Genetic interactions are functional dependencies between genes, which become apparent when the phenotypic effect of one mutation is altered by the presence of a second. In model organisms such as yeast, genetic interactions can be rapidly assessed through the systematic construction of double mutants and measurement of quantitative phenotypes such as growth rate. Quantitative interactions may be positive or negative, indicating less or more severe double mutant phenotypes than expected from the single mutant phenotypes. Many large genetic network maps have been constructed from high-throughput genetic interaction screens in yeast, providing insight into the global landscape of interactions within the cell as well as the functional relationships between specific components of biological processes and pathways [1–5].

Recently, we used genetic interaction mapping in a "differential mode" to compare the changes in genetic networks across experimental conditions [6–8]. To demonstrate this approach, called differential epistasis mapping, we compared the difference between quantitative genetic interaction scores derived from yeast grown on standard versus DNA-damaging media [6]. We found substantial changes in interaction patterns and demonstrated that the difference in scores was more

effective than the scores in either static condition for highlighting interactions relevant to the pathway under study (DNA damage response). Other biological networks, such as protein-protein interaction (PPI) or protein-DNA interaction networks, have also progressed from observing single experimental conditions to comparing the changes in interactions across multiple experimental conditions or genetic backgrounds. For example, Wrana and colleagues developed the LUMIER (luminescence-based mammalian interactome mapping) strategy to identify pairwise PPIs among a set of human factors with and without stimulation by transforming growth factor β [9]. Similarly, Workman et al. used genome-wide chromatin immunoprecipitation to focus on changes in transcription factor binding after exposure to the DNA damaging agent methyl methanesulfonate (MMS) [10]. More recently, a quantitative approach has been presented by Bisson et al. for measuring differential interactions in PPI networks [11]. This approach, which the authors call affinity purification-selected reaction monitoring (AP-SRM), was used to map quantitative changes in interaction with the protein Grb2, which changes showed that the composition of Grb2 complexes was remarkably dependent on the stimulation. By focusing on additional hub proteins beyond Grb2, this method is likely to be useful for obtaining a global overview of protein network remodeling in response to a stimulus.

The progression from static to differential network biology in many fields increases the need for specialized statistical strategies for scoring differential networks. One approach to improving differential signal is to use paired experimental designs that reduce the noise between treated and untreated measurements. For example, experimental designs such as the two-color microarray were originally developed to reduce the noise resulting from technical variability, and various statistical methods have been developed to leverage the paired structure of these experiments (reviewed in [12–15]). Similar to two-color microarrays, differential network measurements can pair treated and untreated measurements. While some of the differential interaction studies [6,7] have employed such an experimental design, they did not utilize this information in their analysis, treating each measurement as independent.

Here, we investigate the statistical structure of two large-scale differential genetic interaction experiments [6,7] and present a generalized strategy for scoring differential genetic interaction data. Our strategy produces differential genetic interaction networks that are more reproducible and more enriched for biologically relevant interactions than previous approaches based on network subtraction. A MATLAB implementation of our strategy is provided as an additional file with the online version of this article.

2.3 Results and discussion

2.3.1 The differential interaction model

The format of a differential genetic interaction experiment takes growth-rate measurements for each double mutant across two or more conditions. A single mutant yeast strain, called the *query*, is mated with an entire set of other single mutants (e.g. deletions of all non-essential yeast genes), referred to as *array* strains. The resulting diploids are sporulated and then undergo multiple selection steps to produce colonies of haploid double deletion mutants. In the last step of the pipeline, the same yeast colonies are replicated onto different media exhibiting the chosen growth conditions (Figure 1a; see [3,6,16] for high-throughput genetic interaction screening protocols).

Because one run of this experimental pipeline produces double mutant colonies that are grown in separate conditions but share the same initial steps, we had reason to believe that the double mutant growth-rate measurements are not independent. Using data from Bandyopadhyay et al. [6], we tested this hypothesis by comparing the correlation of experimental replicates (i.e. colonies generated in separate pipelines but grown in the same condition) with the correlation of colonies generated in the same pipeline but grown in different final conditions. Strikingly, we found that the correlation of colonies grown in different conditions was

much greater than the correlation of experimental replicates (Figure 1b), even though the experimental replicates were grown under identical growth conditions and the conditional replicates were not. This observation suggested some degree of statistical dependence between the conditional replicate measurements.

We further assessed the dependence across the conditional measurements with an analysis of the variance of replicate measurements. Assuming independence, the difference between two normally distributed random variables is distributed normally, with a variance equal to the sum of the variances of the original distributions (Equation 1).

$$\mathcal{N}(\mu_1, \sigma_1^2) - \mathcal{N}(\mu_2, \sigma_2^2) \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2) \quad (1)$$

Therefore, for each double mutant, the variance of the differences between the static measurements should be equal to the sum of the variances of the static measurements.

Using the data from two differential interaction mapping experiments comparing MMS and standard growth conditions [6,7], we found that the variance of the difference for each double mutant was less than half of the expected differential variance, and even less than the variance of static (non-differential) measurements (Figure 2). These results confirm that the across-condition measurements are not

independent and raise the possibility that significant error reduction may be achieved by the differential mode of analysis.

2.3.2 The dS Score: A quantitative measure of differential interaction

Accordingly, we developed a strategy for scoring differential genetic interactions, which accounts for the dependency structure of the data. Assuming a growth constant *p* for each plate, which captures plate-to-plate differences in growth rate, the observed double mutant colony size z_{qai} can be factored as follows:

$$z_{qaic} = p_{qic} \cdot f_{qc} \cdot f_{ac} + \epsilon_{qaic} \tag{2}$$

Where *q* and *a* represent the query and array strains, *i* represents the experimental replicate, *c* represents the condition, *f* indicates the single mutant fitnesses, and ϵ represents the residual. Collins et al. [17] developed a strategy that uses colony size population trends to estimate *p*, f_q, and f_a and obtain a measurement of the residual, which serves to quantify the degree of genetic interaction between the query and array mutants.

For differential interactions, the null or "non-interaction" model is that the mean of the differences between paired residuals is equal to zero:

$$\sum_{i}^{n} \frac{\epsilon_{qaic} - \epsilon_{qaic_0}}{n} = \sum_{i}^{n} \frac{\delta_{qaic}}{n} = 0 \tag{3}$$

where *c* indicates the treatment and c_0 indicates the untreated, or reference, condition, and δ represents the difference in colony size residuals. Assuming the ϵ are normally distributed, the degree to which this mean differs from zero given the variance of the replicates can be modeled using the paired t-statistic. We call our statistic the *dS score*, "d" for "differential" and "S score" after the name of the statistic used by Collins et al. [17].

$$\mathrm{dS\ score} = \frac{\bar{\delta}_{qac}}{s_{qac}/\sqrt{n}} \tag{4}$$

where δ_{qac} is the mean of the differences of the residuals (Equation 3) and s_{qac} is the sample standard deviation of the differences of the residuals. Unlike the S-score [17], we found that the sample variance was the best approximation of the variance (based on the quality control metrics described below) and did not employ a minimum bound or any modifiers or priors (such as in the case of SAM, Cyber-T, or LIMMA in microarray analysis [18,19,15], see also [20]).

2.3.3 Similarity of differential interaction profiles provides distinct functional information

Previously, it has been shown that the correlation of static interaction profiles identifies many gene functional relationships not identified by direct genetic interactions (a genetic interaction profile is the

set of all interactions with a given gene) [1,17]. Given our new quantitative score for differential interactions, we therefore investigated whether differential interaction profiles could also be used to provide distinct functional information. Indeed, we found that the correlation of differential interaction profiles was able to identify relationships relevant to the treatment response and, furthermore, that these links were not identified either by direct interactions (static or differential) or by correlation of static profiles.

For example, using the dS score, we observed a very high differential similarity score between SWI4 and the subunits of the HIR complex (Figure 3). In contrast, when computing genetic profile similarity between SWI4 and HIR in either static condition (standard or MMS-treated), similarity scores were strikingly low. SWI4 is the DNA-binding member of the SBF complex, a key regulator of genes involved in DNA synthesis and repair in G1 to S phase [21,22]. HIR1, HIR2, and HIR3 are subunits of the HIR complex which negatively regulate histone protein transcription [23] under control of the DNA-damage checkpoint kinase DUN1 [24]. Although SWI4 and HIR have not been previously implicated in a genetic relationship, SWI4 has been shown to regulate histone gene expression [25,26] suggesting that an interaction between SWI4 and HIR is feasible, especially in context of the DDR. Thus, differential similarity can

identify functional relationships between genes that are not apparent from profile similarity analysis in static conditions.

We identified a total of 99 functional associations like SWI4 and HIR, i.e. gene pairs with low static similarity and high differential similarity. These gene pairs indicate DDR-relevant interactions that would not be identified through previously available methods. One of the key limitations of static profile similarity is that the static profile is populated by interactions pertaining to both the treatment as well as general cell growth. These non-relevant interactions diminish the similarity between genes that otherwise function very similarly in the treatment response. Additionally, the larger variance inherent in the static measurements contributes to noisier interaction profiles which decreases the similarity of otherwise related profiles. Differential interactions are effective at identifying treatment-relevant relationships because they cut down the noise and eliminate non-related interactions.

2.3.4 Performance of the dS score and differential profile similarity

We investigated the quality of the dS score by examining its false discovery rate, reproducibility and biological enrichment. As a baseline for comparison, where applicable dS scores were compared to the differential p-values described by Bandyopadhyay et al. [6], which indicate an empirically-determined significance for the difference in S

scores between two conditions. We designate the –log p-values from Bandyopadhyay et al. [6] as the "B score". To estimate the false discovery rate of different dS score thresholds, we first generated a dS null distribution using the data from Bandyopadhyay et al. [6], in which the final step involved pinning each double mutant twice in the same condition. These two colonies were paired and scored as if they were colonies grown in separate conditions (corresponding to z_{qaic} and z_{qaic0} in eqn. (2) above). We observed that the dS score has approximately symmetric false discovery rates for positive and negative scores (Figure 4a).

Next, we assessed reproducibility of the dS score by comparing B and dS scores generated using replicates 1-3 and, separately, 4-6 from Guénolé et al. [7]. Using only gene pairs that were scored in both analyses, we found that the dS score yields a much tighter reproducibility across replicates than the B score (Figure 4b-c; Additional file 4, Figure S1). In particular, the Pearson correlation across replicates was remarkably higher for the dS score than the B score (Figure 4d; the values on the far right correspond to data shown in Figure 4b-c). We found it of particular interest that for the most significant interactions, the dS score tends to greater and greater reproducibility, while the reproducibility of the B score

drops to zero, indicating that for larger and larger values, the B score picks up on less and less signal.

To measure the biological enrichment of the dS score, we generated a bronze-standard set of interactions similar to that used by Bandyopadhyay et al. [6]. We included in our standard set any gene pair in which both genes were annotated as "DNA-damage response" (DDR) in the Gene Ontology [27] (corresponding to 903 or 2,575 gene pairs in the Bandyopadhyay et al. [6] or Guénolé et al. [7] data sets, respectively), as well as any gene pair defined by the YeastNet 2.0 benchmark set [28] containing at least one DDR gene (390 or 772 gene pairs, respectively). As a second standard, we used the set of co-complex interactions compiled by Baryshnikova et al. [29], which is based on the set of macromolecular complexes recorded in the Saccharomyces Genome Database (www.yeastgenome.org) or in the CYC2008 protein complex catalogue [30]. Using these two standards, we generated precision-recall plots for two previously-published differential interaction networks (Bandyopadhyay et al. [6] and Guénolé et al. [7]). This analysis indicated that the dS score has essentially the same precision for recovering the DDR and the co-complex standards as the original p-values published by Bandyopadhyay et al. [6](Figure 5; see also Additional File 4, Figures S2-4). However, we observed a notable improvement in enrichment for DDR

interactions when using profile similarity of dS scores as compared to profile similarity of B scores (Figure 5a,b).

Additionally, it is well known that gene pairs with high profile similarity are often members of the same physical complexes [31,32], so we investigated whether the same is true for differential-profile similarity. We found that the genes with similar dS score profiles are strikingly more enriched for co-complex pairs (Figure 5c,d), and specifically for protein complexes involved in the DDR (Additional file 4, Figure S2). For example, differential profile similarity was able to achieve a precision of 60-100% for recovering either DDR pathway interactions or protein complexes, using data from either of two studies. This performance was in contrast to that of individual differential interactions, which had a precision of 1-20% using these same standards and data.

It is interesting that B score profile similarity is under-enriched for meaningful relationships. Part of this behavior may be explained by our observation that extreme B score values tend to capture noise and are not reproducible (Figure 4b-d). Because profile similarity is heavily influenced by larger values, B score profile similarity is overly sensitive to noise. Thus, relatively few spurious interactions can have an extensive influence on profile similarity.

We finally compared dS scores and dS profile similarity scores to the static S scores and profile similarity scores from the same data. We found that differential similarity scores are more enriched for DDR interactions than static similarity scores, even though static scores are more enriched for non-DDR specific interactions (Additional file 4, Figure S3).

The reasons for the improved performance in identifying relevant genetic relationships of the dS score over the B score and the static scores deserve some attention. Genetic interaction mapping experiments are subject to many systematic sources of noise. For example, the ratio of double mutant cells to single mutant cells in the colonies growing on the single-mutant selection plate (see Figure 1 for an outline of the experimental workflow) affects the observed double mutant fitness in the following step. Other sources of systematic noise include uneven agar surfaces, which affect the quantity of material that is picked up and deposited during plate pinning, and variations in incubation time, humidity, etc. (Table 1). Despite sophisticated data processing methods, traces of these systematic artifacts may be preserved, and this noise can influence the estimation of interaction effects. The current experimental design for static interaction mapping experiments does not control for these artifacts, and the previous method for scoring differential interactions did not take advantage of built-in controls. However, our

approach uses the paired relationships between plates to eliminate many sources of systematic noise, increasing our ability to identify reproducible and relevant differential interactions (Figures 1, 2, 4). This result is of broad interest because finding the appropriate control plays an important part in differential experimental design in many fields.

2.3.5 Interpretation of the dS score

The previous approach to scoring differential interactions derived a score from the difference between static interaction scores in each condition. This explicit comparison of scores led to a natural discussion about the interpretation the differential score based on the sign and magnitudes of the static scores [6]. However, because the dS score is not based on the difference between static scores, we suggest the dS score be interpreted following the same logic as static interaction scores. In the static case, positive interactions generally denote gene relationships within the same pathway or complex, while negative interactions generally indicate gene relationships that span parallel or redundant pathways [33]. The difference between differential and static interpretation is that static scores indicate interactions that affect general cell growth, whereas differential scores indicate interactions that affect the treatment response.

While the theoretical interpretation of the dS score is straightforward, the practical interpretation is more complex because the static interaction scores provide a context for the interpretation of the dS score. For example, a gene pair exhibiting a positive interaction in untreated conditions that is more positive in MMS (yielding a positive dS score) should be interpreted differently than an interaction that is negative in untreated conditions that becomes positive in MMS (also yielding a positive dS score). According to the standard interaction model, the latter example is supposedly going from a between-pathway relationship in untreated conditions to a within-pathway relationship in the treatment, which quality the former example does not have, even though both examples exhibit a co-pathway relationship in the DDR response. These various classes of differential interactions exhibit different enrichment rates for our DDR standard (Additional file 4, Figure S4), suggesting that there may be unique qualities to each class, but a more detailed investigation of differential interaction interpretation is left for future work.

2.4 Conclusions

Here, we have put forth a quantitative differential interaction score, the dS score, based on important statistical information inherent in the experimental design. This score not only provides more information about

each interaction than previous approaches, but also shows improved reproducibility and comparable biological enrichment. Additionally, quantitative differential interactions give rise to differential interaction profiles, which we demonstrate to be biologically relevant and uniquely insightful. Furthermore, we provide a new interpretation for differential interactions based on the accepted interpretation of static genetic interactions. We conclude that our differential interaction score is preferred to the previous approach for differential genetic interaction mapping analysis.

2.5 Methods

2.5.1 Correlation of query replicates

We used normalized colony size residuals to calculate the correlation of query replicates (Figure 1b). Our approach to computing these residuals is based on the approach published by Collins et al. [17]. In brief, the raw colony sizes are pre-processed to filter bad colonies and correct spatial artifacts. Each plate (i.e. the set of all colony sizes from the same plate) is normalized by the plate mode, calculated using a kernel density estimation method [34]. Next, array single mutant fitnesses are estimated using the median normalized colony size for a given array position across all plates, which are then subtracted from the respective double mutant colony sizes to yield normalized colony size residuals. These

residuals are, in turn, used to calculate several quantities: (1) the pair-wise correlation for each pair of conditional plate replicates, i.e. double mutant selection plates derived from the same single mutant selection plate differing only in the growth condition; (2) pairwise correlation of untreated experimental replicates; and (3) pairwise correlation of randomly selected queries.

2.5.2 The dS score

Normalized differentials are obtained by subtracting untreated normalized colony sizes from the corresponding treated normalized colony sizes. The dS score is then computed as the pooled t-statistic of the six replicates for a given double mutant versus all double mutant measurements containing the respective array gene deletion. Note that the S score, for scoring static interactions, employs a minimum bound on the variance of the six double mutant replicates [17], while the dS score does not bound the variance.

2.5.3 Scoring null differential interactions

The null distribution of dS scores was generated by using replicate pairs of measurements grown on the same plate (and therefore same condition) and following the same scoring procedure already described. The differentials for the three replicates in each condition were pooled to produce six total replicates for each gene pair. We computed false

discovery rates for each dS score cutoff as the ratio of the proportion of null scores beyond the cutoff to the proportion of observed dS scores beyond the cutoff.

2.5.4 Biological enrichment

The "bronze" standard for differential genetic interactions in response to DNA damage was compiled as (1) the set of all gene pairs in which both genes are annotated as "DNA damage response" (DDR) in the Gene Ontology [27] (term ID GO:0006974, direct association; accessed December 2011), and (2) the set of all gene pairs indicated by the YeastNet 2.0 benchmark set [28] in which at least one gene is annotated as DDR. The lists of DDR genes and bronze-standard DDR gene pairs are provided as additional files.

The gold standard used for co-complex membership is defined by Baryshnikova et al. [29]. Precision-recall plots were computed using the absolute value of the dS scores (treating positive and negative interactions equally).

2.5.5 Significance of Pearson correlation

To assess the significance of the difference between the correlation coefficients of the scores in Figure 3, we calculated the correlation of bootstrapped data for 10,000 iterations in a paired fashion and counted

the number of cases in which the correlation of B scores was greater than the correlation of the dS scores.

2.5.6 Determining associations similar to SWI4-HIR

To identify gene associations similar to SWI4 and HIR, where the differential similarity is high and the static similarity is low, we used the cutoffs of >0.35 and <0.15 for differential and static similarity scores, respectively.

2.6 List of abbreviations

DDR – DNA-damage response

SGA – Synthetic Genetic Array

2.7 Competing interests

The authors declare that they have no competing interests.

2.8 Authors' contributions

GB developed the statistical model and performed the validation and discovery. Both authors read and approved the final manuscript.

2.9 Additional data files

The following additional data are available with the online version of this paper. Additional data file 1 is a table of the dS, S, and profile similarity scores for the data from Bandyopadhyay et al. [6]. Additional

data file 2 is a table indicating the gene pairs used as the DNA damage response bronze standard in our study. Additional data file 3 is the MATLAB implementation of our method. Additional data file 4 is a PDF containing our additional notes and figures.

2.10 Acknowledgements

The authors thank Rohith Srivas, Koyel Mitra, and Philip Jaeger for providing insight and advice. This work was supported by National Institutes of Health grants ES014811 and GM084279 to TI and 5T32GM008666-13 to Bruce Hamilton.

Chapter 1, in full, is a reprint of the material as it appears in Differential analysis of high-throughput quantitative genetic interaction data. Bean, G. J. & Ideker, T. Genome Biology. 13, R123 (2012). The dissertation author was the primary investigator and author of this paper.

2.11 Figures

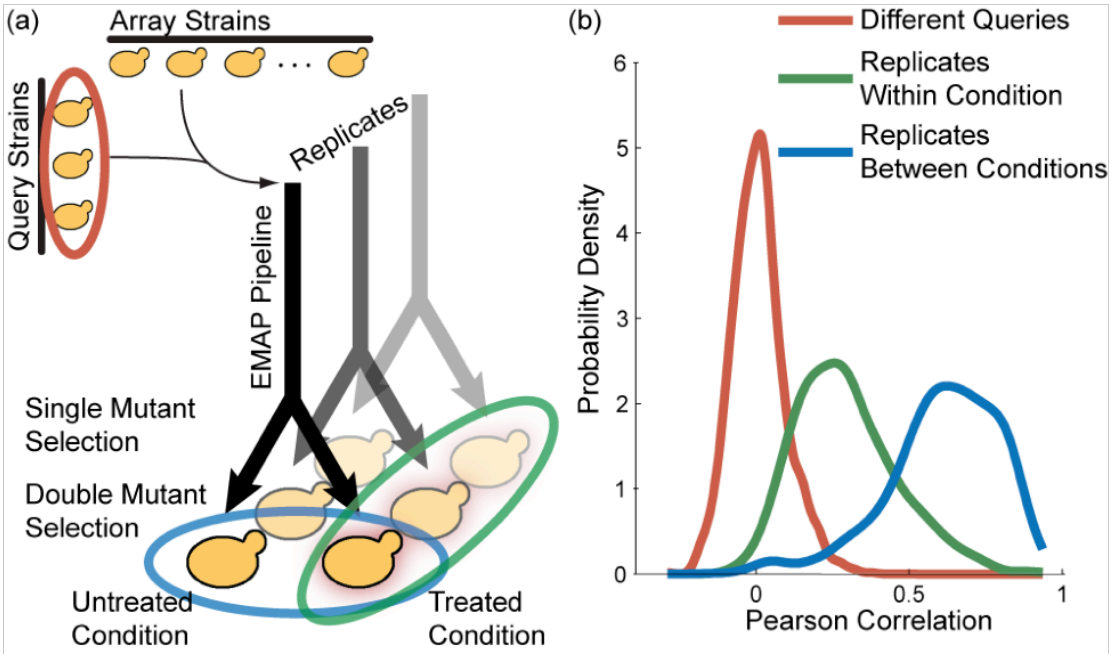


Figure 2.1 The paired experimental pipeline.
(a) The pipeline for generating differential genetic interactions is the same as for static genetic interactions except for a split onto treated and untreated plates in the last step. **(b)** Normalized colony size profiles for the same experimental replicate across the two conditions (blue) have the greatest Pearson correlation, as compared to the profiles of two experimental replicates of the same condition (green) or the profiles of different queries (red).

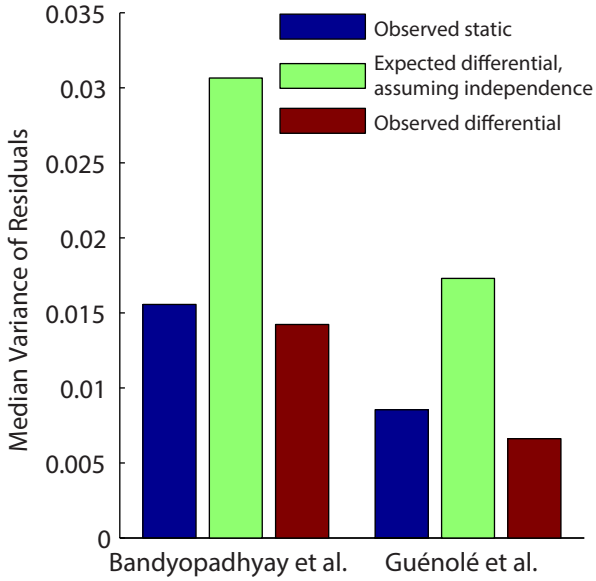


Figure 2.2 Theoretical and observed differential variances.
Bar plot of the observed static, expected differential (assuming independence), and observed differential variances of normalized colony size residuals. The median values across all double mutants are shown.

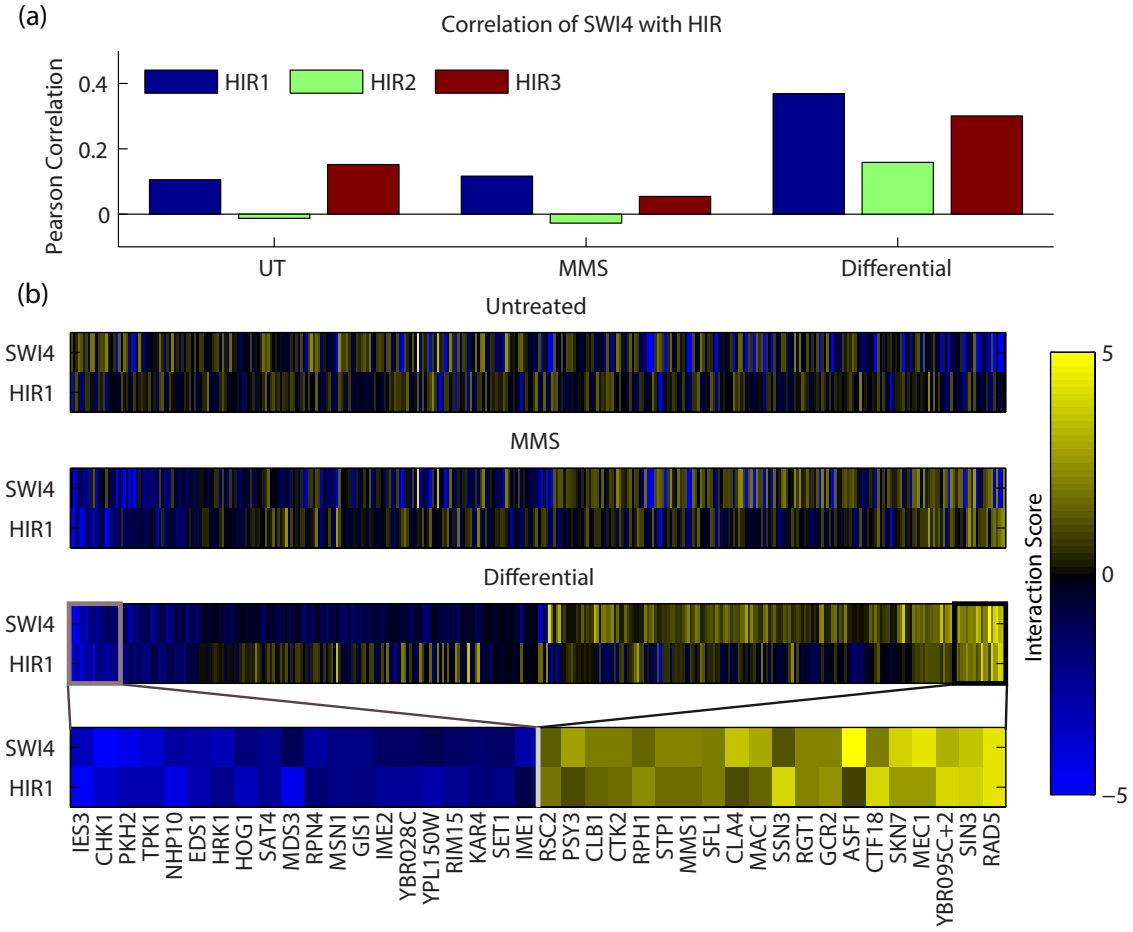


Figure 2.3 Differential profile similarity between SWI4 and HIR.
(a) Bar plot showing the Pearson correlation of HIR1/2/3 profiles with SWI4 for untreated, MMS, and differential (dS) scores. **(b)** Heatmaps of the untreated, MMS, and differential interaction profiles of SWI4 and HIR1; the bottom panel illustrates the interactions with greatest similarity between SWI4 and HIR1.

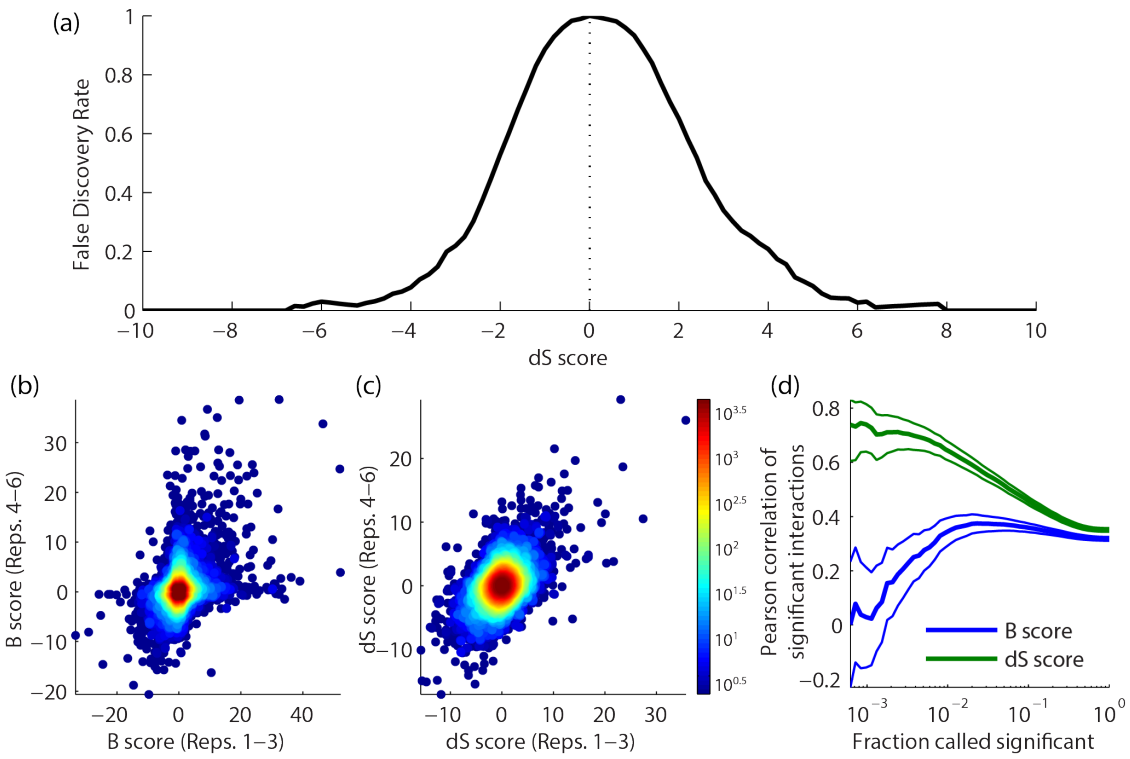


Figure 2.4 False discovery rate and reproducibility the dS score.
(a) Plot of the false discovery rate of the dS score as a function of score magnitude. **(b-c)** Scatter of differential scores calculated on independent replicate subsets using (b) the B scores and (c) the dS score; the points shown in either panel are only those scored by both analyses. **(d)** Plot comparing the Pearson correlation of significant interactions for the B and dS scores (blue and green, respectively) over a full range of significance thresholds – i.e. the correlation of the top *n* percent of the interactions for n = 0.1% (left side) to n = 100% (right side); error bars indicate the 95% confidence intervals of the correlation coefficient.

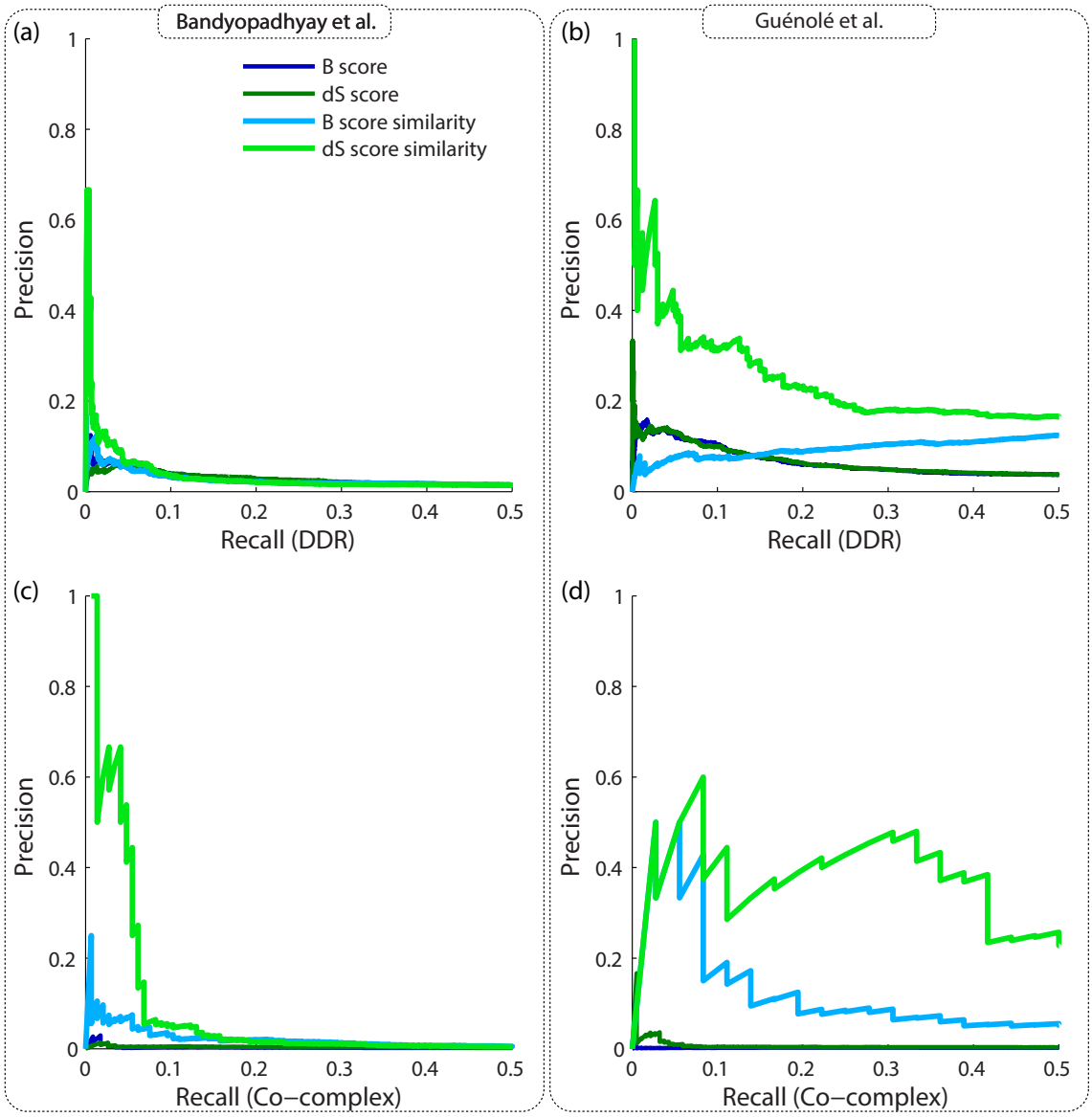


Figure 2.5 Performance of dS score and differential profile similarity. **(a-d)** Precision-recall plots comparing the biological enrichment of B and dS scores and their corresponding profile similarity scores for DDR interactions (a,c) and co-complex interactions (b,d) using the data from Bandyopadhyay et al. (a,b) and Guénolé et al. (c,d).

Table 2.1 Sources of noise and their effect on interaction scores.

Source of noise	Noise affects score?	
	Static score	dS score
Double/single mutant ratio, pre-DM selection	✓	
Double/single mutant ratio, DM selection	✓	✓
Uneven agar surface, pre-DM selection	✓	
Uneven agar surface, DM selection	✓	✓
Variation in environment, pre-DM selection	✓	
Variation in environment, DM selection	✓	✓

2.12 References

1. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S, Prinz J, St Onge RP, vandersluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin Z-Y, Liang W, Marback M, Paw J, San Luis B-J, Shuteriqi E, Tong AHY, van Dyk N, Wallace IM, Whitney J, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C: The genetic landscape of a cell. *Science (New York, N.Y.)* 2010, 327:425–3110.1126/science.1180823.

2. Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, Silva AC, Shales M, Collins SR, van Wageningen S, Kemmeren P, Holstege FCP, Weissman JS, Keogh M-C, Koller D, Shokat KM, Krogan NJ: Functional organization of the S. Cerevisiae phosphorylation network. *Cell* 2009, 136:952–6310.1016/j.cell.2008.12.039.

3. Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, Weissman JS, Krogan NJ: Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 2005, 123:507–1910.1016/j.cell.2005.08.031.

4. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, Schuldiner M, Gebbia M, Recht J, Shales M, Ding H, Xu H, Han J, Ingvarsdottir K, Cheng B, Andrews B, Boone C, Berger SL, Hieter P, Zhang Z, Brown GW, Ingles CJ, Emili A, Allis CD, Toczyski DP, Weissman JS, Greenblatt JF, Krogan NJ: Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 2007, 446:806–1010.1038/nature05649.

5. Zheng J, Benschop JJ, Shales M, Kemmeren P, Greenblatt J, Cagney G, Holstege F, Li H, Krogan NJ: Epistatic relationships reveal the functional organization of yeast transcription factors. *Molecular systems biology* 2010, 6:42010.1038/msb.2010.77.

6. Bandyopadhyay S, Mehta M, Kuo D, Sung M-K, Chuang R, Jaehnig EJ,

Bodenmiller B, Licon K, Copeland W, Shales M, Fiedler D, Dutkowski J, Guénolé A, van Attikum H, Shokat KM, Kolodner RD, Huh W-K, Aebersold R, Keogh M-C, Krogan NJ, Ideker T: Rewiring of genetic networks in response to DNA damage. *Science (New York, N.Y.)* 2010, 330:1385–910.1126/science.1195618.

7. Guénolé A, Srivas R, Vreeken K, Wang S, Krogan NJ, Ideker T, van Attikum H: Dissection of DNA Damage Response Pathways using a Multi-Conditional Genetic Interaction Map. *Molecular Cell* 2013. In press.

8. Ideker T, Krogan NJ: Differential network biology. *Molecular Systems Biology* 2012, 8:1–910.1038/msb.2011.99.

9. Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, Shinjo F, Liu Y, Dembowy J, Taylor IW, Luga V, Przulj N, Robinson M, Suzuki H, Hayashizaki Y, Jurisica I, Wrana JL: High-throughput mapping of a dynamic signaling network in mammalian cells. *Science (New York, N.Y.)* 2005, 307:1621–510.1126/science.1105776.

10. Workman CT, Mak HC, mccuine S, Tagne J-B, Agarwal M, Ozier O, Begley TJ, Samson LD, Ideker T: A systems approach to mapping DNA damage response pathways. *Science (New York, N.Y.)* 2006, 312:1054–910.1126/science.1122088.

11. Bisson N, James DA, Ivosev G, Tate S a, Bonner R, Taylor L, Pawson T: Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nature biotechnology* 2011, 29:653–810.1038/nbt.1905.

12. Patterson T a, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu T-M, Bao W, Fang H, Kawasaki ES, Hager J, Tikhonova IR, Walker SJ, Zhang L, Hurban P, de Longueville F, Fuscoe JC, Tong W, Shi L, Wolfinger RD: Performance comparison of one-color and two-color platforms within the microarray Quality Control (MAQC) project. *Nature biotechnology* 2006, 24:1140–5010.1038/nbt1242.

13. Cui X, Churchill G a: Statistical tests for differential expression in cdna microarray experiments. *Genome biology* 2003, 4:210.

14. Cui X, Hwang JTG, Qiu J, Blades NJ, Churchill G a: Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics (Oxford, England)* 2005, 6:59–7510.1093/biostatistics/kxh018.

15. Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* 2004, 3:Article310.2202/1544-6115.1027.

16. Tong a H, Evangelista M, Parsons a B, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C: Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science (New York, N.Y.)* 2001, 294:2364–810.1126/science.1065810.

17. Collins SR, Schuldiner M, Krogan NJ, Weissman JS: A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome biology* 2006, 7:R6310.1186/gb-2006-7-7-r63.

18. Tusher VG, Tibshirani R, Chu G: Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* 2001, 98:5116–2110.1073/pnas.091062498.

19. Baldi P, Long AD: A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics* 2001, 17:509–519.

20. Murie C, Woody O, Lee AY, Nadon R: Comparison of small n statistical tests of differential expression applied to microarrays. *BMC bioinformatics* 2009, 10:4510.1186/1471-2105-10-45.

21. Sidorova J, Breeden L: Analysis of the SWI4 / SWI6 Protein Complex , Which Directs G1 / S-Specific Transcription in Saccharomyces cerevisiae. 1993, 1310.1128/MCB.13.2.1069.Updated.

22. Ho Y, Mason S, Kobayashi R, Hoekstra M, Andrews B: Role of the casein kinase I isoform, Hrr25, and the cell cycle-regulatory transcription factor, SBF, in the transcriptional response to DNA damage in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* 1997, 94:581–6.

23. Spector M, Raff A, desilva H: Hir1p and Hir2p function as transcriptional corepressors to regulate histone gene transcription in the Saccharomyces cerevisiae cell cycle. *Molecular and cellular* 1997, .

24. Sharp J a, Rizki G, Kaufman PD: Regulation of histone deposition proteins Asf1/Hir1 by multiple DNA damage checkpoint kinases in Saccharomyces cerevisiae. *Genetics* 2005, 171:885–9910.1534/genetics.105.044719.

25. Kato M, Hata N, Banerjee N, Futcher B, Zhang MQ: Identifying combinatorial regulation of transcription factors and binding motifs. *Genome biology* 2004, 5:R5610.1186/gb-2004-5-8-r56.

26. Eriksson PR, Ganguli D, Clark DJ: Spt10 and Swi4 control the timing of histone H2A/H2B gene activation in budding yeast. *Molecular and cellular biology* 2011, 31:557–7210.1128/MCB.00909-10.

27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 2000, 25:25–910.1038/75556.

28. Lee I, Li Z, Marcotte EM: An improved, bias-reduced probabilistic functional gene network of baker's yeast, Saccharomyces cerevisiae. *Plos one* 2007, 2:e98810.1371/journal.pone.0000988.

29. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, Youn J-Y, Ou J, San Luis B-J, Bandyopadhyay S, Hibbs M, Hess D, Gingras A-C, Bader GD, Troyanskaya OG, Brown GW, Andrews B, Boone C, Myers CL: Quantitative analysis of fitness and

genetic interactions in yeast on a genome scale. *Nature methods* 2010, 7:1017–2410.1038/nmeth.1534.

30. Pu S, Wong J, Turner B, Cho E, Wodak SJ: Up-to-date catalogues of yeast protein complexes. *Nucleic acids research* 2009, 37:825–3110.1093/nar/gkn1005.

31. Srivas R, Hannum G, Ruscheinski J, Ono K, Wang P-L, Smoot M, Ideker T: Assembling global maps of cellular function through integrative analysis of physical and genetic networks. *Nature protocols* 2011, 6:1308–2310.1038/nprot.2011.368.

32. Bandyopadhyay S, Kelley R, Krogan NJ, Ideker T: Functional maps of protein complexes from quantitative genetic interaction data. *Plos computational biology* 2008, 4:e100006510.1371/journal.pcbi.1000065.

33. Boone C, Bussey H, Andrews BJ: Exploring genetic interactions and networks with yeast. *Nature reviews. Genetics* 2007, 8:437–4910.1038/nrg2085.

34. Parzen E: On estimation of a probability density function and mode. *The annals of mathematical statistics* 1962.

CHAPTER 3

DEVELOPMENT OF ULTRA-HIGH-DENSITY SCREENING TOOLS FOR MICROBIAL "OMICS"

Gordon J. Bean, Philipp A. Jaeger, Sondra Bahr, and Trey Ideker.

3.1 Abstract

High-throughput genetic screens in model microbial organisms are a primary means of interrogating biological systems. In numerous cases, such screens have identified the genes that underlie a particular phenotype or a set of gene-gene, gene-environment or protein-protein interactions, which are then used to construct highly informative network maps for biological research. However, the potential test space of genes, proteins, or interactions is typically much larger than current screening systems can address. To push the limits of screening technology, we developed an ultra-high-density, 6144-colony arraying system and analysis toolbox. Using budding yeast as a benchmark, we find that these tools boost genetic screening throughput 4-fold and yield significant cost and time reductions at quality levels equal to or better than current methods. Thus, the new ultra-high-density screening tools enable researchers to significantly increase the size and scope of their genetic screens.

3.2 Introduction

Large-scale genetic screening experiments (i.e. simultaneous analysis of many mutants, either pooled or arrayed) have enabled researchers to identify gene functions and functional relationships underlying many processes (for numerous examples see *Nature Reviews Genetics* series "The Art and Design of Genetic Screens"). In an increasing number of model organisms, such screens take advantage of available mutant libraries, including complete collections of gene knockout strains, over-expression constructs, and the like. In the most typical mode, these screens identify genes that are required for, or modulate, a phenotype of interest. Very similar screens can be performed to identify gene and protein interactions using systems such as Synthetic genetic arrays (SGA) and Yeast two-hybrid (Y2H)[1-5].

In this context, single-cell organisms have proven extraordinarily useful due to their ease of genetic manipulation and straightforward growth conditions. Suitable species can be found in bacteria (e.g. *E. coli*), fungi (e.g. *S. cerevisiae, S. pombe)*, and algae (*C. reinhartdii*), allowing researchers to assess the effects of a gene across large evolutionary timescales [6-9]. However, even the comparatively small genomes of these model species contain thousands of genes that can be screened in

any number of growth conditions. Furthermore, screening for combinations of mutants, such as in genetic interaction screening or physical interaction mapping, requires hundreds of thousands to millions of possible strains.

Most current screening methods rely on growing microbial model organisms on a solid, nutrient-rich agar surface in a regular grid pattern to allow for reliable parallel quantification of a simple phenotype such as growth [6-8]. The maximum density of microbial colonies per surface unit, the duration of necessary growing time, and the sensitivity and robustness of the downstream image acquisition and analysis pipeline are all important factors that determine screening throughput, and present screening systems typically allow for up to 1536 colonies per agar plate [10].

Here, using *S. cerevisiae* as a benchmark, we substantially enhance screening throughput by enabling growth and analysis of 6144 mutant yeast colonies on a single agar plate. The significance of achieving this number is that the vast majority of microbial model organisms have gene counts very near but not exceeding this number, allowing for an entire, genome-wide screen to be performed on a single agar plate. We evaluate data quality and cost performance of this new, ultra-high-

density colony-transfer system in comparison to current methods, and provide a free computational toolset for ultra-high-density image analysis.

3.3 Material and methods

3.3.1 6144-density pad development

The new 6144-density pads were produced in collaboration between Singer Instrument Co. Ltd. (Roadwater/UK), KREO Technologies (Oakville, ON/Canada) and S.B. in the Boone laboratory (University Toronto, ON/Canada). Different pressure molds were cut and several trial pads were cast by varying plastic temperature, injection pressure etc. Pads were evaluated for flatness, stiffness, and pinhead quality using standard, SingerPlus plates as well as aluminum and polytetrafluoroethylene SingerPlusPlate+ prototypes. The pad with the overall best performance and tolerance was chosen for production and is now commercially available (Singer Instrument Co. Ltd.).

3.3.2 Yeast deletion strains, agar plates, and media preparation

The yeast strains used in this report are based on the commercially available yeast knockout (YKO) strain collection (Thermo Fisher Scientific Inc., Waltham/MA) with kanamycin as a deletion marker. The collection was stored in glycerol stocks at -80°C in 96-well format until used. We produced higher-density plates by first pinning thawed glycerol stocks

onto agar plates and then robotically combining 96 plates into increasingly higher densities. Media and agar plates were composed following established E-MAP protocols [6-8,10-12] in standard Singer plate clones (IGENE Supplies, Shanghai/China). Microtiter-format agar plates were poured manually with 42ml of liquid agar-medium in each plate, cooled on the bench top overnight, and were allowed to dry for 24hrs at room temperature. It is critical for high-density pinning that the agar surface is dry before pinning starts.

3.3.3 1536- and 6144-density-format pinning

To achieve estimates of technical and biological variances, we pinned a minimum of 18 replicate plates of each format and imaged each plate at 0, 3, 6, 9, 12, 24, and 48 hours after pinning (the 96 and 384 plates were not imaged on the 3- and 9-hour time points). Additionally, we imaged every plate's source plate immediately before pinning. Overall, we acquired over 1200 high-resolution plate images. Unless otherwise specified, measurements and analyses using the 6144 format were done on the 12-hour images, while measurements and analyses using the 384- and 1536 formats were done using the 48-hour images.

All liquid-to-solid and solid-to-solid yeast transfers were conducted using a Singer RoToR robotic plate handler (Singer Instrument Co. Ltd). 96- to 1536-format pinnings were performed using the respective factory

standard settings for source and target plates. 1536-to-6144 (1536x4) pinnings with a 1536 pad were performed with default factory settings at the source plate (with 0.15 mm offset) and custom settings for the target plate (pin pressure 64%, speed 10mm/s, overshoot 1mm, no offset). 6144-to-6144 pinnings with a 6144 pad were performed with custom settings at the source (pin pressure 50%, speed 10mm/s, overshoot 0.6mm, no offset) and the target plate (pin pressure 64%, speed 10mm/s, overshoot 0.6mm, no offset).

For clean 1536-to-6144, 6144-to-6144, and 6144-to-24576 transfers it is essential not to overgrow the source plates (max. 6-12hrs incubation) as the pinheads are very small and overly large source colonies lead to cross-contamination and smear formation between neighboring colonies. For hyper-density plates, we first pinned 6144 source plates, incubated those for only 3hrs at room temperature and the immediately pinned again to 24576.

3.3.4 Digital image acquisition

All digital images were acquired with a commercially available SLR camera (18Mpixel Rebel T3i, Canon USA Inc., Melville/NY) with an 18-55mm zoom lens. We used a white diffusor box with bilateral illumination and an overhead mount for the camera in a dark room. Images were taken in highest quality, 8-bit JPEG. Down-sampling experiments suggest

that 10Mpixel cameras should be sufficient for 6144-format image acquisition (data not shown).

3.3.5 Image analysis and data processing

Images were normalized, spatially corrected, and quantified using a set of custom algorithms (aka "The Colony Analyzer Toolkit") written in Matlab (MathWorks Inc., Natick/MA). The complete software package is available online (https://github.com/brazilbean/Matlab-Colony-Analyzer-Toolkit). The workflow for measuring colony size from digital images is as follows: (1) the image is cropped to the plate, (2) the colony grid is overlaid on the plate image, (3) the size of each colony is measured, (4) colony sizes are plate normalized, and (5) colony sizes are spatially corrected to remove local, nutrient-based growth effects. Colony size quantification and analysis generally followed the same procedures used in other studies [6,11].. However, the toolkit can be easily adapted to measure other features besides size, such as color or average pixel intensity. An additional version of the toolkit, written in Python, is under development and will be made available at https://github.com/brazilbean/Python-Colony-Analyzer-Toolkit.

3.4 Results

3.4.1 Technical and computational improvements required for ultra-high-density plates

Current transfer pads only exist up to a maximum density of 1536 pins per pad. To increase colony density further, we developed a novel transfer pad with 6144 individual pinheads, allowing us to print plates with 6144 or 24576 clearly defined yeast colonies (Material and Methods). To compare the various density performances, we grew single gene deletion mutants from the haploid yeast knockout strain collection in 384-, 1536-, and 6144-colony grids in large plate replicate numbers for 48hrs (N=18). Mutant colonies grown at lower densities were a subset of those grown at higher densities, allowing us to compare the same 384 yeast mutant fitness values across densities (Figure 1). Colony sizes were measured and analyzed as described in the Material and Methods.

Growing 6144 yeast colonies on a single agar plate introduces unique challenges in plate image analysis, rendering previous data extraction approaches inadequate [6,10-13], and led us to develop a new image analysis and data normalization software package (available at https://github.com/brazilbean/Matlab-Colony-Analyzer-Toolkit). Small errors in orientation in the 6144 format are sufficient to misalign colonies at the ends of each row or column due to the increased proximity of colonies

(Figure 2a). Consequently, we needed an algorithm that could precisely identify the locations of each colony in the 6144 grid, which is particularly challenging with small colonies (i.e. 0-hour time point, Figure 2b). We achieved a highly-accurate grid alignment by estimating the angle of grid orientation using the aspect ratio of the cropped image (Figure 2c and Supplemental Note 1 included with the original online version of this chapter), estimating the locations of the four corners of the grid based on grid dimensions and spacing, and iteratively interpolating the remaining locations of the grid/colony positions by describing each location as a linear function of the grid row and column positions [6,13] (Figure 2d).

Previous algorithms [6] relied on a single plate-wide pixel-intensity threshold to measure colony sizes, an approach we found to be inaccurate leading to over- and under-estimation of small colony sizes. This effect was particularly obvious at early time points with small colonies (0-3 hrs 'Global thresholding' Figure 3a,b) and at late time points for peripheral colonies (24-48 hrs 'Global thresholding' Figure 3b). We overcame this problem by implementing an algorithm that determines a dynamic pixel-intensity cutoff for each colony based on the distribution of local background pixel-intensities (Figure 3c). Additionally, when colonies are overgrown, the pixels between adjacent colonies may have intensities greater than the threshold, resulting in multiple colonies being counted as

one, a problem we solved by determining the local minima surrounding each colony (Figure 3d). In summary, we developed an image-processing pipeline with an improved ability to correctly identify colonies over a large dynamic range of colony sizes, using sensitive local background detection (Figure 3e and 4) and improved colony identification algorithms.

3.4.2 Growth performance, cost, and signal quality of ultra-high-density plates

Utilizing our image analysis software package, we then compared the growth and noise performance of 1536-colony and 6144-colony formats (referred to as *1536* and *6144* formats). While typical screens in the 1536 format are imaged 48 hours after pinning, we found that plates in the 6144 format begin to overgrow after 12 hours (Figure 5a); therefore, we make all our comparisons between the 1536/48-hour images and the 6144/12-hour images. Despite the decreased incubation time, 6144 plates exhibit a comparable fold-growth to 1536 plates (~7-fold versus ~8-fold, Figure 5b), because the much smaller pin heads on the 6144 pads produce smaller starting colony sizes (Figure 6a-c).

To benchmark the usability of our new technology, we performed colony-based fitness measurements on yeast gene-deletion mutants, identical to the technique used in epistasis or chemo-genetic screens. The goal of any fitness-based genetic screen is to establish which mutants

exhibit a significant deviation from expected fitness levels. The fitness variance across replicates is a crucial parameter that determines what minimal fitness difference ("fitness resolution") can be called significant. We observed that the variance of colony measurements increased as plate density increased, and decreased as colonies were grown for a longer time (Figure 6d). Consequently, the 6144 format suffers from lower fitness resolution than the 1536 format, assuming identical numbers of replicates (1536 and 6144 N=6, Figure 5c). However, the 6144 format's higher density allows more replicates to be run at equal or lower cost, which can increase 6144 fitness resolution to be as good or even better than the 1536 fitness resolution (6144 N=12, 18, 24, Figure 5c).

Additionally, the dynamic range of fitness levels was slightly larger in 6144 compared to 1536 plates (Figure 6e), indicating that 6144 plates could be cost and quality competitive. To explore the relationship between cost (i.e. replicate number) and fitness resolution, we calculated the minimum fitness phenotype resolvable as a function of the number of replicates and found 6144 plates to be less expensive at equal quality (17% cost reduction). Conversely, better quality could be achieved at equal cost (+7% fitness resolution) compared to 6 replicates of 1536 plates (Figure 5d). We also calculated the fraction of single mutants found significantly different from wildtype at a specific replicate number or cost

level ($p<0.05$, Figure 5e) and observed that the 6144 format can be used to increase the percentage of mutants with a significant phenotype at equal cost (+2.5%), or to allow for cost savings at equal identification level (25% cost reduction, Figure 5e). Importantly, 6144 plates allow the experimenter to dramatically decrease expenses without risking statistical errors that arise with small numbers of replicates. For example, a typical experiment might run 6 replicates in the 1536 format, or 18 replicates in the 6144 format; however, reducing the number of replicates by a factor of 3, 2 replicates in the 1536 format would produce data subject to statistical errors, while 6 replicates of the 6144 plates would not, causing only a mild degradation in data quality. In summary, 6144 plates can be used to improve data quality at equal cost, to cut cost by about 15-20% while maintaining equal data quality, or to finely choose the quality-for-cost balance for a given experiment, while substantially reducing screening time in all cases (-75%, 12hrs vs. 48hrs).

3.4.3 Ultra-high-density plates from high-density pads

Higher-density plates can be produced from lower density source plates by pinning multiple source plates onto a common target plate, a process we call "up-scaling" (Figure 7a). We examined whether it is possible to use 1536 pads to create 6144 plates with the same beneficial properties as described above. We found that 6144 plates pinned with

1536 pads ("1536x4") and analyzed *via* the normal data analysis pipeline showed a dramatic increase in variance (Figure 7b, "1536x4 w/o intra-plate correction"). Further analysis revealed that most of this variance increase was due to batch effects unique to each source plate that were preserved in the target plate (Figure 7c, top). Analyzing the colonies pertaining to each source plate as separate, lower-density plates improved the variance dramatically (Figure 7c, bottom, and b,"1536x4"). Overall, up-scaling using 1536-pads yields growth-curves, dynamic ranges, and quality levels close to 6144 pad pinned plates (Figure 8). However, four pinning pads are used to up-scale for each ultra-high-density plate, thus cost benefits cannot be realized using up-scaling (Figure 7d, assuming a cost ratio of 2:1 for plates to pads and equal cost per pad).

3.4.4 Feasibility of hyper-density plates with 24576 individual colonies

Here we have described the technical qualities of the emerging 6144 format we developed for yeast high-throughput screening, and one could certainly consider the possibility of future increases in colony-density. Given the promising findings on data quality using up-scaling, we explored the technical feasibility of pinning 24576 yeast colonies onto a single agar plate using 6144 pads ("hyper-density plates"). Using a modified 6144 pinning protocol, we successfully created plates with this tremendous colony density, allowing us to run a genome-wide single

mutant screen with four replicates on a single agar plate (Figure 7e). At such small scales the spatial precision required to place pinheads in perfect geometric alignment (to avoid uneven colony neighbor distances) increases. However, we found colony alignment to err by less than one pixel on average (Figure 7f). The resulting fitness values, while increasingly noisy, nevertheless correlate well with lower density data (Figure 7g). At these densities, colonies are only 100-200μm in diameter and new technical challenges arise: colony images become pixelated, resulting in more noise, and the contrast between colony and background pixel intensity decreases, making colony detection more difficult. Furthermore, colonies arrayed at this density begin to overgrow within 3 to 6 hours, reducing the amount of time for growth-defects to be manifest. However, we believe that this hyper-density format could find useful screening applications, especially when combined with further improvements in data acquisition and analysis.

3.5 Discussion

Using a heavily utilized genetic screening system for benchmarking, our analysis suggests that the new 6144 format can be used to effectively cut experimental costs and duration while still maintaining the same level of discriminatory power and quality as the old 1536 format. The new format also gives the researcher more flexibility to choose the desired

cost-quality balance with dramatic cost savings possible (>50%). Our study of the feasibility of hyper-density plates with 24576 individual yeast colonies shows that simultaneous fourfold-coverage of the whole yeast genome on a single agar plate at reasonable quality-levels is now technologically possible.

Our study suggests that our ultra-high-density pad and analysis pipeline exceeds the technical standards necessary to produce high quality data in different types of screens that are based on the quantification of individual microbial colonies. While recent developments in genetic screening technology have enabled researchers to perform some genetic screening experiments in pooled, liquid form [14,15] many "omics" screening technologies have or cannot be adapted to a this type of format, and these screens will undoubtedly benefit from the availability of the ultra-high-density pads.

Intriguingly, our results suggest that if extreme data quality were desired and cost irrelevant, one could pin low-density plates using higher-density pads (e.g. only every 4^{th} or every 16^{th} colony pinned), allowing for the benefits of very large fold-increases of growth (due to small pin heads) to be combined with longer incubation times leading to colony sizes with small biological and technical noise. These considerations and other

experimental techniques may be useful for further extending the limits of high-throughput screening technology.

Data quality could also be improved through modifications to the image acquisition system. As the purpose of this study was to compare 1536- and 6144-format experiments, we did not conduct thorough investigations into the relative merits of image acquisition improvements. However, adjustments in the number of megapixels, lens types, and aperture settings could be made to accomplish incremental improvements in image and data quality. Similarly, researchers desiring to use the 6144-format pads should still be able to acquire data of reasonable quality even if some element of their imaging system is not the same as ours (e.g. their camera provides 12Mpixels instead of 18Mpixels).

3.6 Glossary

epistasis – in double-mutant genetic screens, epistasis occurs when the phenotype of the double mutant deviates from the predicted combination of the single mutant phenotypes; epistasis is indicative of functional relationships among the genes perturbed by the mutations.

fitness – in genetic screens, fitness refers to the ratio of the mutant phenotype (i.e. quantitative measurement, typically growth rate) to the control phenotype; e.g. a mutant with a fitness of 0.9 grows at 90% the growth rate of the control strain.

fitness difference – the difference in fitness between strains

fitness variance – the variance of repeated measurements of a strain's fitness

fitness resolution – the minimum fitness difference needed to determine with statistical confidence that two strains exhibit different fitness phenotypes; greater dynamic range among possible fitness values and smaller fitness variance improve fitness resolution.

hyper-density – 24576 colony format

ultra-high density – 6144 colony format

3.7 Funding

This work was supported by NIH grants GM084279 and GM085764 to T.I. and a grant to develop a "High density replicator System" by Genomics Ontario Inc. to S.B. This research was conducted while P.A.J. was an Ellison Medical Foundation/AFAR Postdoctoral Fellow.

3.8 Author contributions

G.J.B., P.A.J., and T.I. conceived the methods, P.A.J. designed the experiments and conducted the high-throughput screening, S.B. performed the pad optimization experiments, G.J.B. and P.A.J. performed the plate imaging, and G.J.B. wrote the analysis software and performed the data analysis. G.J.B., P.A.J. and T.I. analyzed the results and wrote the manuscript.

3.9 Acknowledgements

We thank D. Jacobsen for his contribution to the pilot experiments for this study, A.R. Carvunis and R. Srivas for critically reviewing this manuscript, and H. Singer for providing sample ultra-high-density pads.

Chapter 2, in full, is a reprint of the material as it appears in Development of ultra-high-density screening tools for microbial "omics". Bean, G. J., Jaeger, P. a, Bahr, S. & Ideker, T. PLoS One 9, e85177 (2014). The dissertation author was the primary investigator and author of this paper.

3.10 Figures

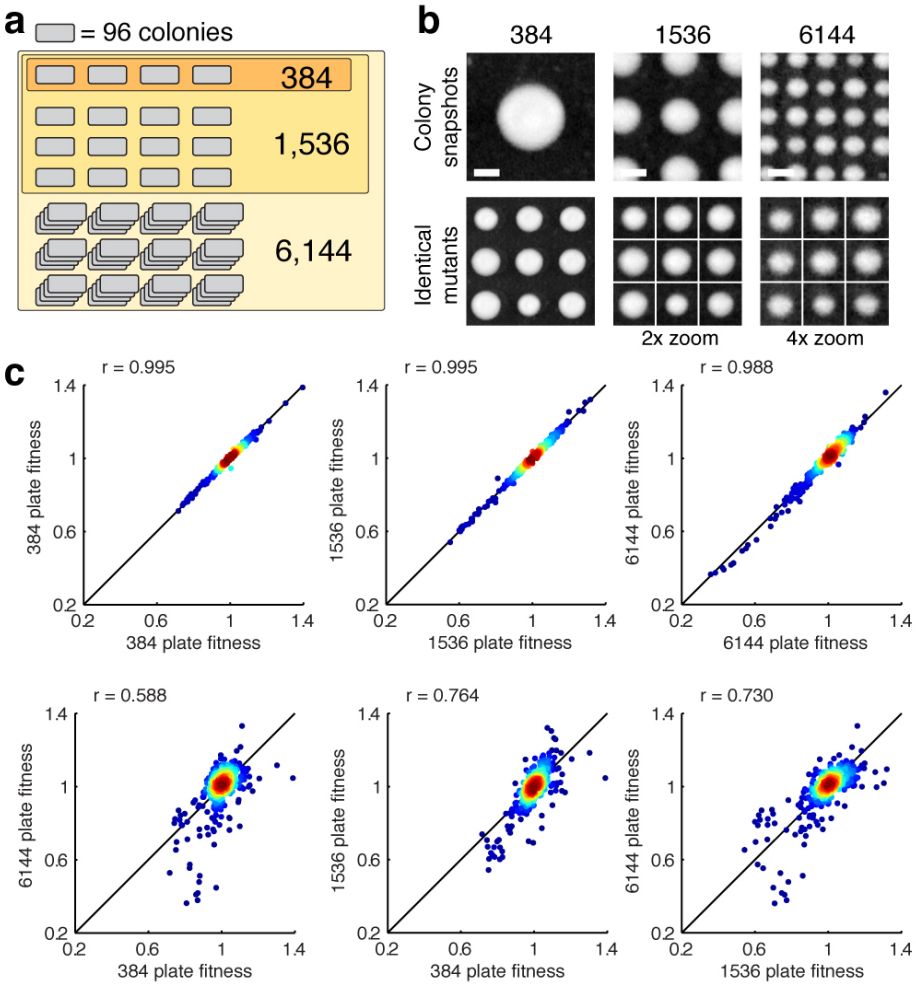


Figure 3.1 Experimental design and correlations between different colony densities.
(a) Diagram indicating the overlap in yeast strains used in the various-colony formats – mutants in lower density plates are always included in higher-density plates. **(b)** Snapshots of 384-, 1536-, and 6144-colony plates (top row, scale bar 1 mm) and mosaic view of identical mutant colonies assembled from 1536- or 6144-colony plates (zoomed) and compared to the 384-colony plate growth (bottom row). **(c)** To compare data quality between formats, we correlated replicates of the different formats internally or across (same format median of N=9 replicates each; across format median of N=18 replicates each).

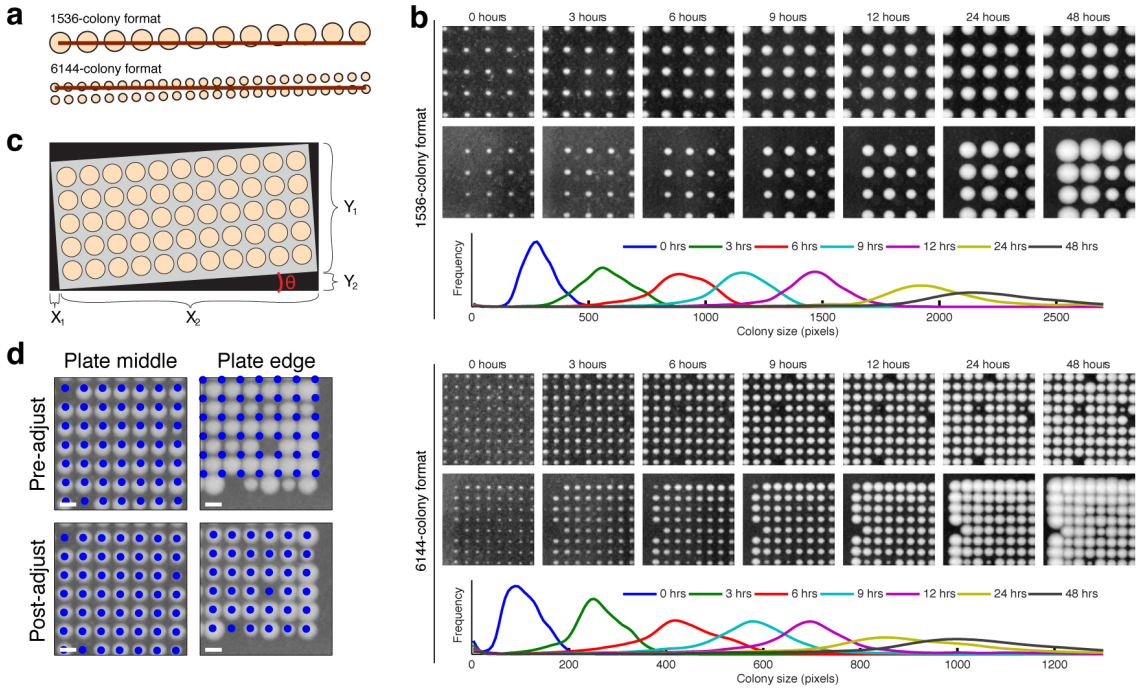


Figure 3.2 Colony growth kinetics and colony grid alignment.
(a) Diagram of rows of larger and smaller colonies, each angled at 0.5 degrees relative to the reference (horizontal bars). Small errors in image rotation in the 6144-colony plates can lead to substantial colony identification errors. **(b)** Time-lapse imaging of the current 1536-density (above) and the new super-high-density format (below) reveals optimal imaging time points of 24–48 hrs for the 1536 and 12–24 hrs for the 6144 format (identical scale for all images). **(c)** Geometric solution for the image rotation problem. Given that the corners of the plate touch the edges of the cropped image, the width and height of the image can each be decomposed into the sum of two smaller values. These four values (X1,X2,Y1,Y2) are all trigonometric functions of h, the angle of orientation of the grid, and the width and height of the plate. These functional relationships comprise a non-linear system of equations with a closed-form solution, which we solved for h. **(d)** Snapshots of colonies growing in the middle and on the edge of 6144-colony plates; blue dots indicate the positions of the grid before (upper row) and after (lower row) the grid-adjustment step (scale bars 1 mm).

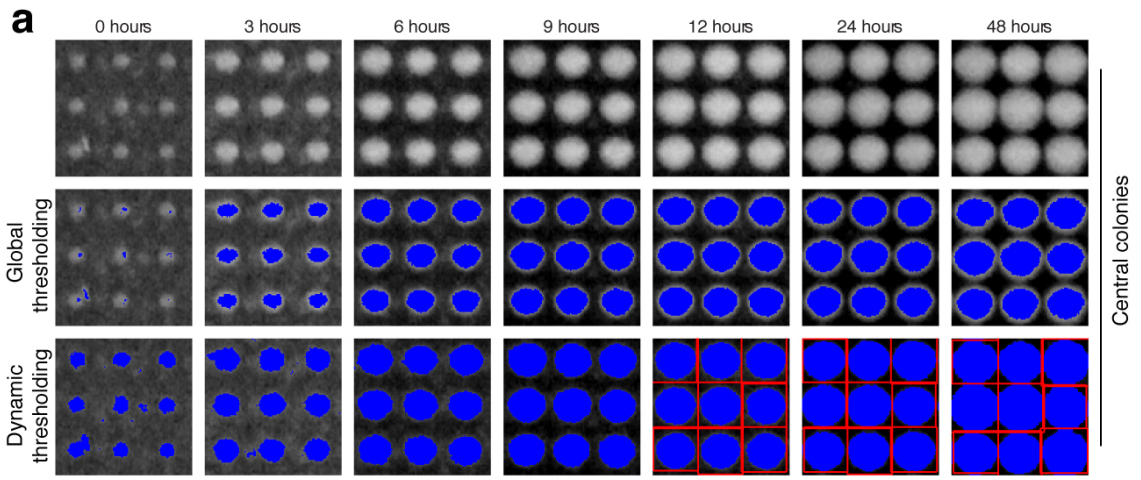


Figure 3.3 Comparison of global and dynamic intensity threshold algorithms.
(a, b) Snapshots of colonies in the plate center and periphery, respectively, 0, 3, 6, 9, 12, 24, and 48 hours; blue pixels in the middle and bottom rows of each panel indicate the pixels called by the respective algorithm as foreground (i.e. colony as opposed to background); red lines indicate predicted colony boundaries. The global intensity used in (a) was computed on the peripheral window, while the global intensity used in (b) was computed on the central window, highlighting the problems of global tresholding. **(c)** Gray-scale snapshots of a single colony at 12 hours (left, scale bar 500 mm); histograms showing the distribution of pixel intensities for the snapshot, the green curve represents the normal distribution fit to the leftmost peak (indicating the distribution of background pixel intensities), blue dotted lines indicate the threshold used to distinguish colony from background (middle); binary output (right, threshold applied). **(d)** Gray scale snapshot centered on an overgrown colony (left, scale bar 500 mm); line plot of median pixel intensity across the center of the snapshot (middle, blue line indicates local intensity threshold, red line indicates the colony boundary); binary image with intensity threshold and bounding box applied (right). **(e)** Local dynamic background estimation is very sensitive and allows for accurate colony-size estimations across a large background intensity range. Original photo of a 0-hour 6144 plate (top); grey scale heat map of the estimated background intensity for each grid position in the 0-hour image (middle; red boxes indicate the positions of the central and peripheral snapshots shown in (a,b)); the reflection of the camera used (bottom) is clearly captured by the background intensity estimation algorithm, demonstrating its sensitivity.

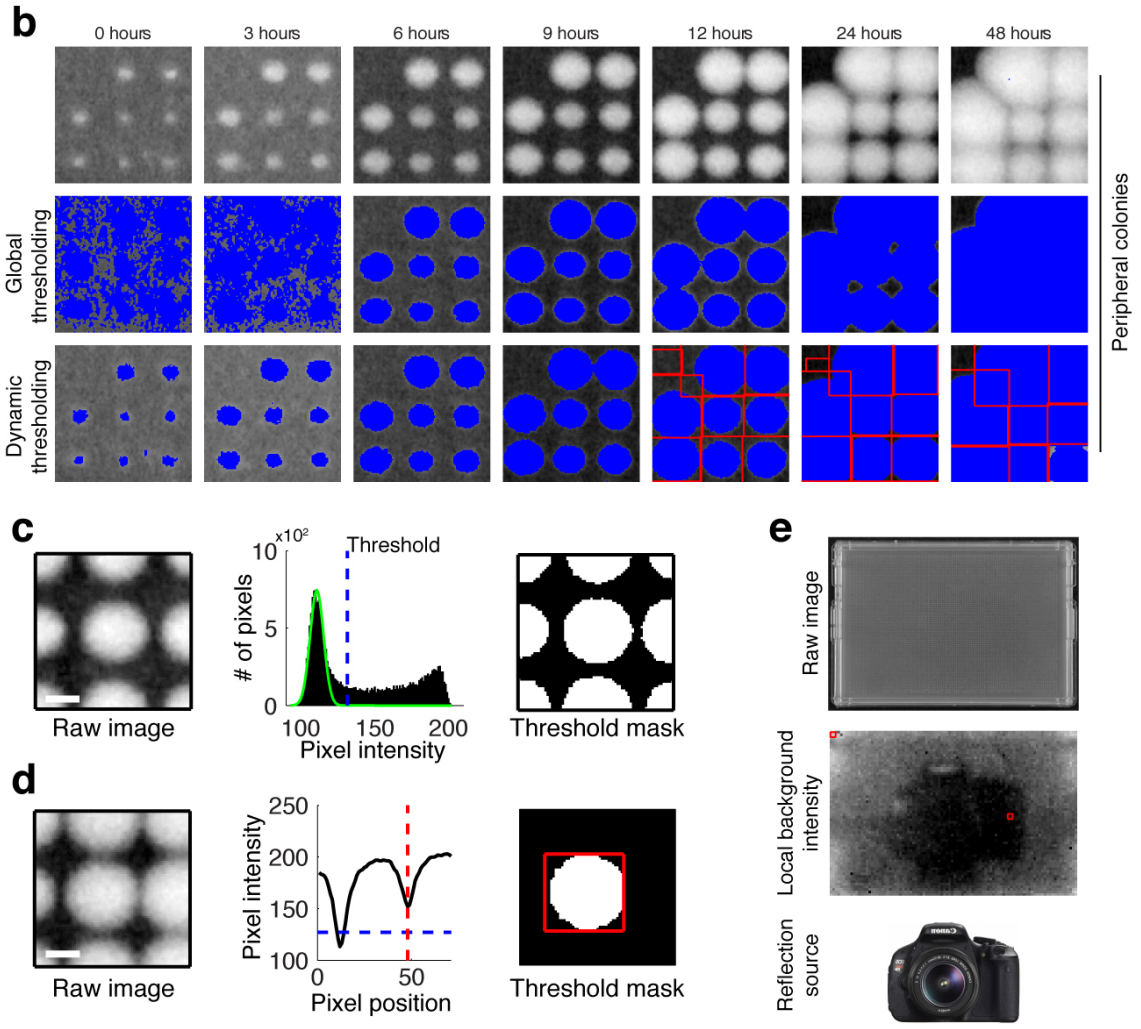


Figure 3.3 Comparison of global and dynamic intensity threshold algorithms. Continued.

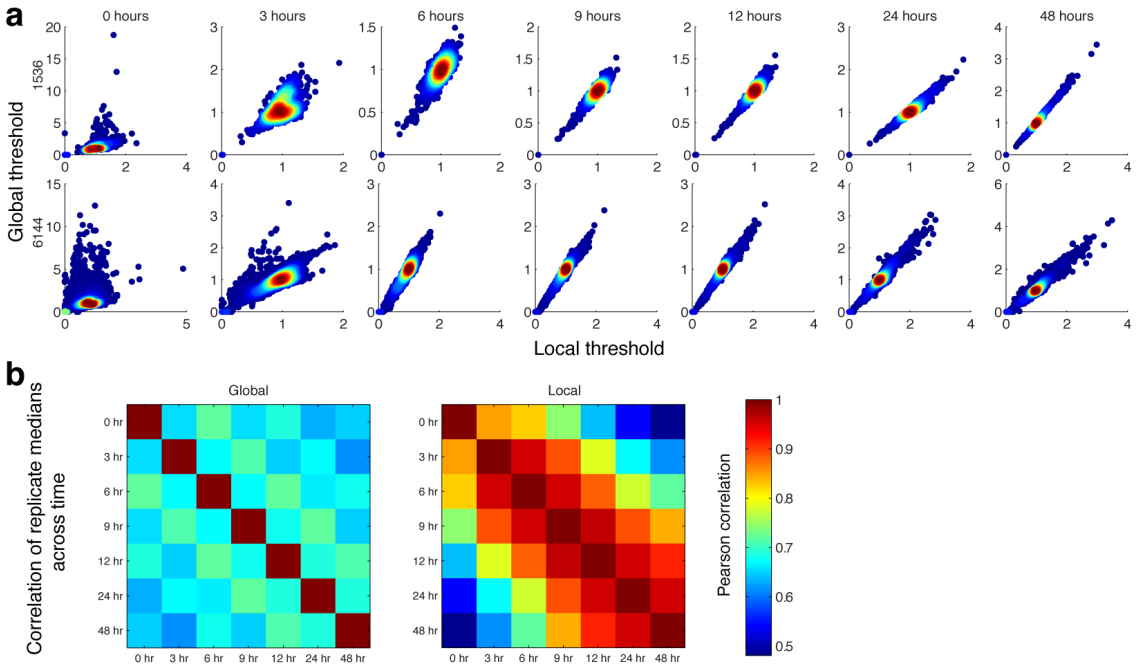


Figure 3.4 Effect of global versus dynamic background.
(a) Comparison of globally or locally/dynamically thresholded colony sizes. While good correlation is achievable between 6 and 24 hrs, poor correlation is observed at the extreme ends of the experiment. **(b)** In general, using dynamic local thresholding (right) achieves much better data correlation across time-points than global thresholding (left).

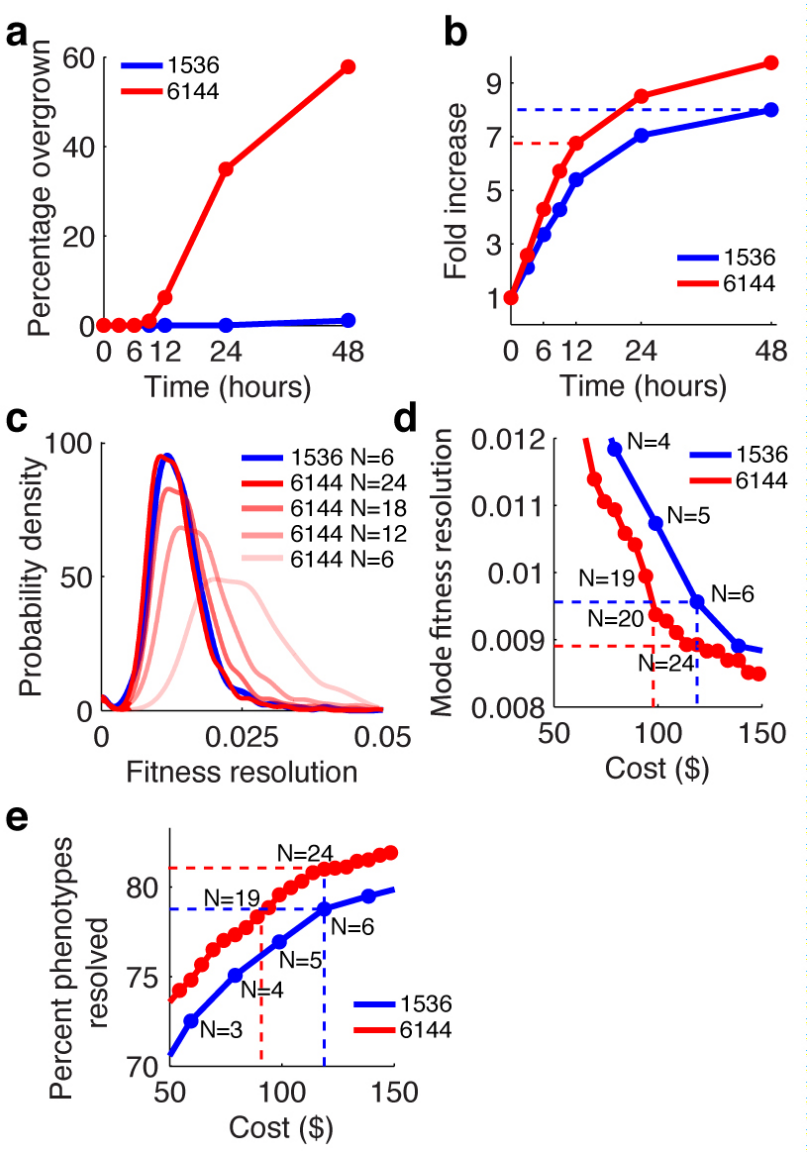


Figure 3.5 Ultra-high-density format data quality and cost efficiency. **(a)** Percentage of colonies overgrown. **(b)** Growth curves based on median colony area fold-increase since pinning (dashed lines indicate fold increase at 12 hrs [6144] or 48 hrs [1536], N=18 for each colony density). **(c)** Distribution of fitness resolutions for 1536 and 6144 format (N= replicate numbers). **(d)** Mode fitness resolutions for a given cost/replicate level (dashed lines indicate equal cost/quality levels, N= replicate numbers). **(e)** Percentage of single mutants that can be resolved (dashed lines indicate equal cost/quality levels, N= replicate numbers).

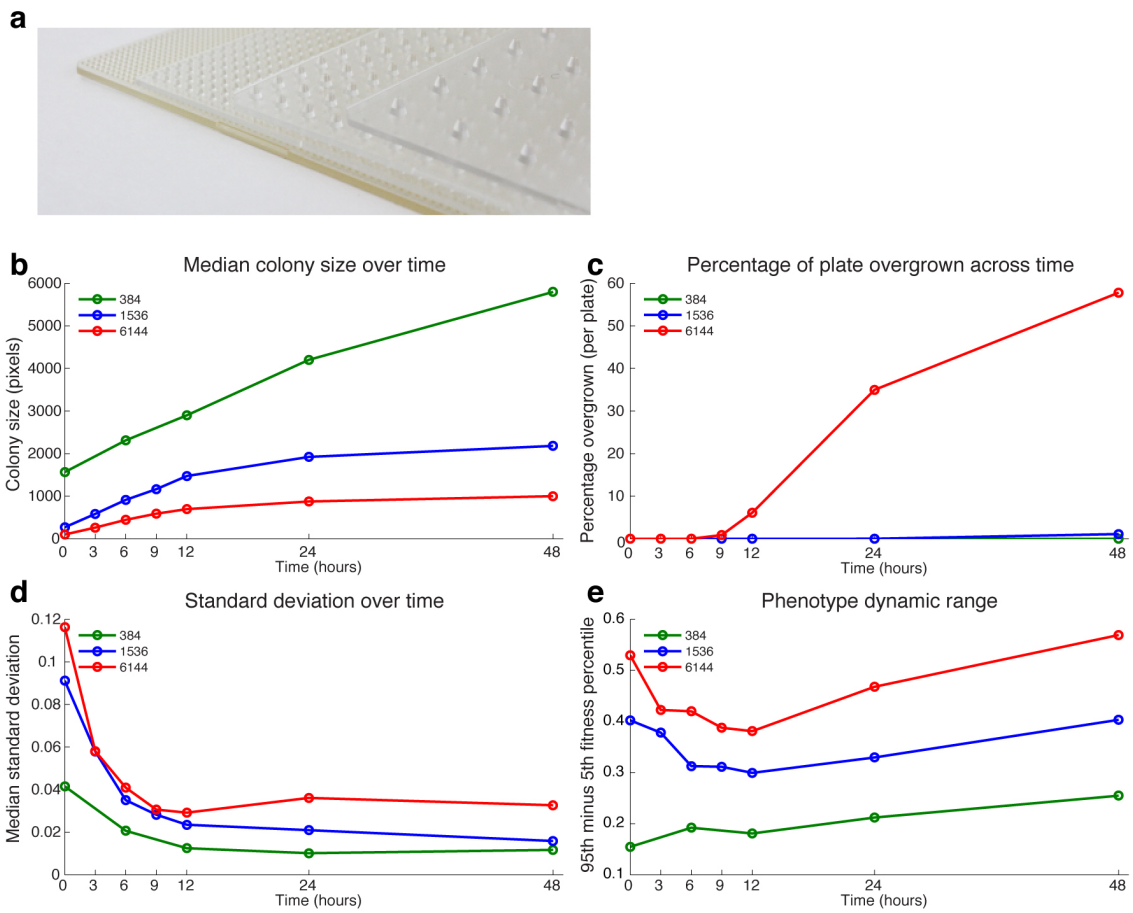


Figure 3.6 Colony size, overgrowth, variance, and dynamic range over time.
(a) Comparison of the different pin pad formats (6144, 1536, 384, 96; left to right). **(b)** Median colony size over time. **(c)** Percentage of plate overgrown over time. **(d)** Standard deviation between replicates over time. **(e)** Phenotype (fitness) dynamic range over time.

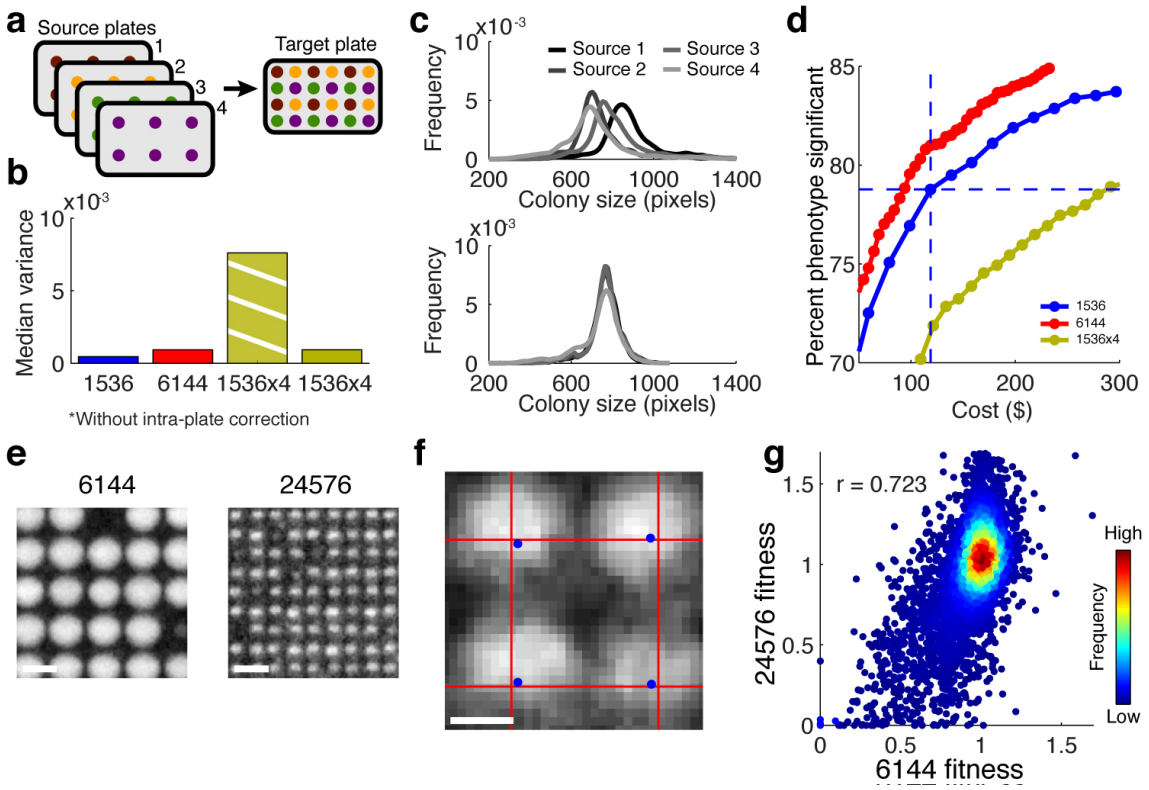


Figure 3.7 Up-scaling and hyper-density.
(a) Schematic of the effects of up-scaling: the combination of different lower-density source plates into one higher-density target plate. **(b)** Comparison of variance in plates pinned with dedicated density-pads (1536, 6144) and plates using up-scaling with or without intra-plate source correction. **(c)** Colony size distributions obtained by the analysis pipeline without (top) and with (bottom) the intra- plate source correction. **(d)** Comparison of the percentage of single mutants that can be identified with a significant fitness phenotype at a given cost/replicate level (--- indicate N=6 at 1536 density). **(e)** Snapshots of 6144- and 24,576 hyper-density colonies at equal scale (scale bar 1 mm). **(f)** Zoomed image showing jitter effect on colony placement; red grid represents perfect alignment, blue dots denote actual pin position (scale bar 100 mm). **(g)** Correlation of fitness measurements obtained with ultra-high- and hyper-density plates.

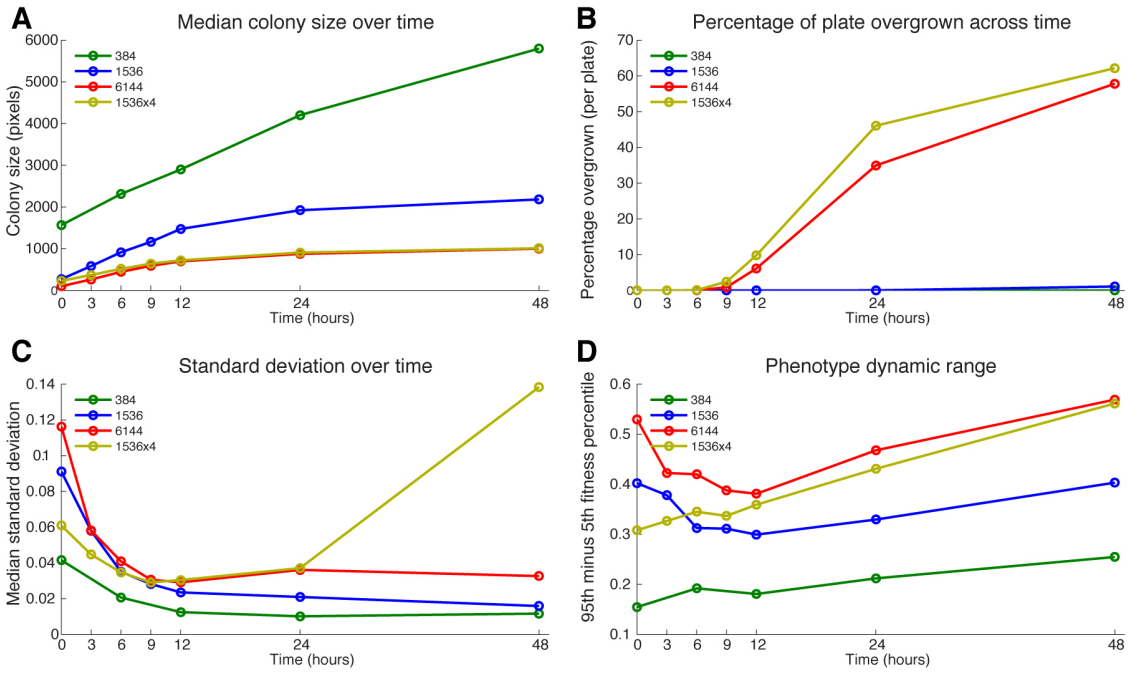


Figure 3.8 Colony size, overgrowth, variance, and dynamic range over time including 1536x4.
(a) Median colony size over time. **(b)** Percentage of plate overgrown over time. **(c)** Standard deviation between replicates over time. **(d)** Phenotype (fitness) dynamic range over time.

3.11 References

1. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403: 623–627. doi:10.1038/35001009.

2. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 98: 4569–4574. doi:10.1073/pnas.061034498.

3. Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins SR, Qu H, Shales M, Park HO, Hayles J, Hoe KL, Kim DU, Ideker T, Grewal SI, Weissman JS, Krogan NJ.. (2008) Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. Science 322: 405–410. doi:10.1126/science.1162609.

4. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Tong AH, van Dyk N,

Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C. (2010) The genetic landscape of a cell. Science 327: 425–431. doi:10.1126/science.1180823.

5. Bandyopadhyay S, Mehta M, Kuo D, Sung MK, Chuang R, Jaehnig EJ, Bodenmiller B, Licon K, Copeland W, Shales M, Fiedler D, Dutkowski J, Guénolé A, van Attikum H, Shokat KM, Kolodner RD, Huh WK, Aebersold R, Keogh MC, Krogan NJ, Ideker T. (2010) Rewiring of genetic networks in response to DNA damage. Science 330: 1385–1389. doi:10.1126/science.1195618.

6. Collins SR, Schuldiner M, Krogan NJ, Weissman JS (2006) A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol 7: R63. doi:10.1186/gb-2006-7-7-r63.

7. Schuldiner M, Collins SR, Weissman JS, Krogan NJ (2006) Quantitative genetic analysis in Saccharomyces cerevisiae using epistatic miniarray profiles (E-MAPs) and its application to chromatin functions. Methods 40: 344–352. doi:10.1016/j.ymeth.2006.07.034.

8. Collins SR, Roguev A, Krogan NJ (2010) Quantitative genetic interaction mapping using the E-MAP approach. Meth Enzymol 470: 205–231. doi:10.1016/S0076-6879(10)70009-4.

9. Ryan CJ, Roguev A, Patrick K, Xu J, Jahari H, Tong Z, Beltrao P, Shales M, Qu H, Collins SR, Kliegman JI, Jiang L, Kuo D, Tosti E, Kim HS, Edelmann W, Keogh MC, Greene D, Tang C, Cunningham P, Shokat KM, Cagney G, Svensson JP, Guthrie C, Espenshade PJ, Ideker T, Krogan NJ. (2012) Hierarchical modularity and the evolution of genetic interactomes across species. Mol Cell 46: 691–704. doi:10.1016/j.molcel.2012.05.028.

10. Wagih O, Usaj M, Baryshnikova A, VanderSluis B, Kuzmin E, Costanzo M, Myers CL, Andrews BJ, Boone CM, Parts L. (2013) SGAtools: one-stop analysis and visualization of array-based genetic interaction screens. Nucleic Acids Res 41: W591–W596. doi:10.1093/nar/gkt400.

11. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, Youn JY, Ou J, San Luis BJ, Bandyopadhyay S, Hibbs M, Hess D, Gingras AC, Bader GD, Troyanskaya OG, Brown GW, Andrews B, Boone C, Myers CL. (2010) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat Meth 7: 1017–1024. doi:10.1038/nmeth.1534.

12. Baryshnikova A, Costanzo M, Dixon S, Vizeacoumar FJ, Myers CL, Andrews B, Boone C.. (2010) Synthetic genetic array (SGA) analysis in Saccharomyces cerevisiae and Schizosaccharomyces pombe. Meth Enzymol 470: 145–179. doi:10.1016/S0076-6879(10)70007-0.

13. Bean GJ (2013) Colony Analyzer Toolkit. User Manual: 1–20.

14. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge

RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G. (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 320: 362–365. doi:10.1126/science.1150021.

15. Ho CH, Magtanong L, Barker SL, Gresham D, Nishimura S, Natarajan P, Koh JL, Porter J, Gray CA, Andersen RJ, Giaever G, Nislow C, Andrews B, Botstein D, Graham TR, Yoshida M, Boone C. (2009) A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. Nat Biotechnol 27: 369–377. doi:10.1038/nbt.1534.

CHAPTER 4

GENOME-WIDE CHARACTERIZATION OF DYNAMIC GROWTH PROFILES VIA MASSIVELY PARALLEL TIME-LAPSE IMAGING

Gordon Bean, Philipp Jaeger, and Trey Ideker

4.1 Abstract

Time-lapse analyses of mutant colonies have long been performed on many experimental platforms, however few studies have identified phenotypes other than growth rate. Here we utilize the new 6144-colony agar format to perform genome-wide time-lapse analysis of the yeast knockout collection to quantify dynamic growth phenotypes and identify key biological processes involved in adaptation to metabolic stress through nutrient depletion. We then apply our method to study the dynamic response to UV-radiation and observe that growth profiles recapitulate key biological processes in DNA repair and suggest novel relationships. We also find evidence supporting a co-function of RAD23 and RAD14 in the UV response.

4.2 Introduction

As yeast cells grow on agar, they change their environment, drawing in nutrients and releasing waste. Consequently, the cell must be able to adapt to its changing environment in order to maintain optimal growth. In classic high-throughput screening approaches using agar plates, the cumulative effect of these dynamic changes is measured as the colony size after a fixed period of time. Using this single measurement, genetic interactions have been mapped for many cellular processes [1–5]; however, not all genetic interactions can be observed in the standard growth condition, and it has been shown that by growing the colonies in different conditions, new, highly specific interactions are observed [6–8].

While these agar-based studies have been remarkably successful using only final colony growth, they do not provide much insight into the dynamic behavior of colony growth before the final image is taken. Full growth curves have been measured using other experimental platforms [9–12], but in these platforms the experimental throughput is limited to 96 or 384 strains per run, making genome-wide experiments costly and time-consuming. Additionally, even when the full growth curve information is available, most studies only quantify the growth rate and do not identify other growth features, such as length of lag phase or total growth

efficiency [9,13]. Even when additional features are measured, all growth curve studies to date have been limited to 96- or 384-well formats.

Here we combine time-lapse imaging with the super-high-throughput capacity of the new 6144-colony format for agar plates [14] to measure full growth curve information for all mutants in the yeast knockout collection on multiple conditions and characterize non-standard growth phenotypes.

4.3 Results

Using a modification of genetic interaction-mapping techniques [15,16], we performed a massively parallel, genome-wide growth curve analysis of all non-essential and some essential yeast genes (Supplemental table S1). The colonies were pinned on synthetic complete (SC) and standard (YPAD) media and imaged every 2 minutes for 24 hours, with 12 replicates per condition. The images were analyzed and the data were corrected, smoothed, and filtered using the Matlab Colony Analyzer Toolkit [14], as described in Methods.

We found that regardless of final colony size (“ total growth efficiency”), normalized growth curves exhibit the same basic shape – that is, when normalized for initial and final colony size, most growth curves are nearly identical (Figure 1a-b). This result is particularly interesting – even when the growth rate of a mutant is reduced, the relative changes

in growth rate over time are the same for almost all mutants. Given the consistency of growth curve shape across many mutants, we use this population shape as a reference curve for quantifying the deviant behavior of mutants showing different curve shapes (Figure 1c). We compute the root mean squared error (RMSE) of each growth curve as a measure of total curve deviation to identify colonies that exhibit deviant behavior. The RMSE is by construction positive, but we apply a sign to distinguish curves that are >50% above the reference (which behavior we call "stall") from those that are mostly below the reference (which behavior we call "lag"), and call our signed RMSE "total deviation". Almost all curves that deviate to an appreciable degree fall either all above or all below the reference; we observed no mutants that oscillate significantly around the reference curve. Finally, to verify that growth curve deviation is not simply a recapitulation of final colony size (the standard metric), we scattered total deviation against normalized colony size and observed that total deviation is not a function of normalized colony size (Figure 1d). Thus, total deviation provides novel information about mutant phenotypes.

We performed a genome-wide GO term enrichment for genes with deviant phenotypes in SC, controlling for false discovery rate (see Methods). We found that pathways and processes that are enriched for

deviant genes include lysine biosynthesis (stall), homoserine biosynthesis (stall), purine biosynthesis (stall), ergosterol biosynthesis (stall), vacuolar acidification (lag), tryptophan biosynthesis (lag), small ribosomal subunit assembly (lag), large ribosomal subunit biogenesis (predominantly lag), large ribosomal subunit assembly (both lag and stall), nonfunctional rRNA decay (lag and stall), exosome (lag and stall), and RNA processing (lag and stall) (Figure 2).

The concurrence of stall phenotypes with biosynthetic pathways led us to the hypothesis that in many cases the stall phenotype may be caused by a premature stationary phase caused by a depletion of key nutrients in the media: i.e. because these mutants cannot synthesize their respective nutrients, once those specific nutrients are depleted from the media, the colony can no longer grow. For example, the lysine and homoserine biosynthesis pathways exhibit stall phenotypes when grown on SC; however, when grown on YPAD media, which is contains a greater abundance of essential nutrients, these same mutants do not show stall phenotypes (Figure 3).

A previous study described mutants with prolonged lag phase as those that struggle to adapt to changing environmental conditions [17]. While the phenotype we have termed “lag” is more general than the previously described lag-phase phenotype, our results are consistent with

this interpretation, as we see several processes involved in turnover of cellular components and machinery exhibiting lag behavior, such as ribosome biogenesis and assembly or vacuolar acidification (necessary for proteolysis and autophagy).

Because of its ability to capture dynamic behavior in growth, we applied our time-lapse analysis to study the cellular response to a dynamic UV radiation treatment. 5 hours after pinning onto SC media, we applied 0.015 Joules of UV radiation (see Methods). After analysis of the growth curves, we found that many of the known radiation-sensitive mutants exhibited a change in growth rate after the UV treatment (Figure 4a). Using the topmost radiation-sensitive mutants, we performed a hierarchical clustering of the deviation profiles (i.e. amount a colony is above or below its expected colony size at each point in time; Figure 4b). We observed that the timing of the change in growth rate differs across these mutants and that mutants form functional groups cluster together. This can be seen when genetic interaction similarity data from Costanzo et al. [1] is arranged in the same order as the clustered deviation data – as genes with similar interaction profiles are arranged next to each other, squares of similar values are formed along the diagonal of the heatmap (Figure 4c).

As the primary feature distinguishing among these clusters is the timing of the change in growth rate after UV treatment, this clustering provides information about when each module in the DNA-damage response is needed for DNA repair. Genes in the nucleotide excision repair (NER) pathway are the first to show a response to the UV treatment, followed by the RAD52 epistasis group which handles double stranded break repair, and finally the RAD24 and RAD9 checkpoints and associated. Remarkably, this ordering of response to UV reflects the timing of gene function in the cell cycle: the first response to UV-induced lesions occurs with NER. Lesions that are not corrected lead to stalled replication forks and double stranded breaks. Throughout the cell cycle, DNA-damage checkpoints seek to ensure that the cell does not proceed with growth until the DNA is repaired – errors not corrected during the course of the cell cycle should be caught by the final checkpoint. Errors that persist can lead to failure to complete mitosis and result in cell death.

The clustering of specific gene pairs, successfully recapitulating double mutant fitness profile scores, is remarkable given that these data were acquired using only single mutant growth curves. Given the functional relevance of the clusters formed, these data may provide novel information about gene relationships. For example, RAD23 is known to bind with RAD4 to comprise a stoichiometric complex called NEF2

[18,19]; however, RAD23 clusters next to RAD14, a DNA-damage recognition factor. These genes have been shown to interact physically, forming a complex that helps to recruit TFIIH to the site of DNA damage [20,21]. Furthermore, while these genes do not show genetic profile similarity in untreated conditions (whereas RAD23 and RAD4 show a low level of similarity), their genetic interaction similarity is on par with that between RAD23 and RAD4 when screened in UV-treated conditions [8] (Figure 4d). These results together suggest that the association between RAD23 and RAD14 is of similar importance as the association of RAD23 and RAD4 in NEF2, in the context of UV-induced DNA damage.

4.4 Discussion

Here we provide new phenotypic information for 4947 gene deletion mutants. Besides providing general, pathway-level information, these data highlight interesting biological relationships that can serve as starting points for future hypotheses. For example, the two genes AAH1 and APT1 both operate on adenine as a substrate; however, AAH1 exhibits a lag phenotype while APT1 exhibits a stall phenotype. Or, SSN2, SSN8, and SRB8 – members of the RNA Polymerase II Mediator CDK8 complex – show lag phenotypes in SC; however, SRB7, a member of the RNA Polymerase II Mediator middle module, shows a stall phenotype. These contrasting phenotypes for related processes and components

may shed light on the role these pathways play in the metabolic function of the cell or contribute to cell growth.

We have provided simple explanations for the processes underlying stall and lag phenotypes; however, it is easy to imagine that many cases require a more nuanced interpretation. For example, most of the amino acid biosynthesis pathways enriched for deviant behavior show stall phenotypes. In contrast, the tryptophan biosynthesis pathway exhibits lag phenotypes. This difference in trend suggests a unique role for tryptophan in cell growth compared to other amino acids. Another factor to consider is how colony efficiency (as measured by normalized colony size) can alter the interpretation of deviant phenotypes. For example, TAT2, the tryptophan-specific importer, shows a stall growth curve shape; however, it also grows faster and achieves greater growth efficiency than the population control, suggesting that the stall behavior may not be caused by premature depletion of a specific nutrient (such as in the lysine pathway cases), but because the rapidly growing colony runs into other constraints, such as glucose depletion or competition with neighboring colonies.

It has been shown recently that examining mutant growth on special combinations of carbon and nitrogen sources provides useful metabolic information [12]. Using the super-high-throughput capacity of

the 6144-colony format makes it more feasible to interrogate many different media types, and time-lapse growth analysis can reveal additional insights that simple endpoint or growth-rate analyses cannot, making our platform a good candidate for future media-mutant interaction screens.

Another recent study identified interesting relationships between amino acid concentrations and various cellular components and pathways [22]. Our platform could serve as a useful tool for identifying metabolically relevant genetic interactions. While we have used single-mutant strains for this study, double-mutants can be screened just as easily. With both deviation and growth efficiency phenotypes to quantify, time-lapse analysis is powered to find more genetic relationships than can be found using colony size alone. With the vast assortment of media types and double mutant combinations possible, there are still many interesting experiments that can be done, and our platform is well adapted to make these screens.

4.5 Conclusion

We performed a genome-wide screen of the yeast gene deletion collection using time-lapse analysis. We discovered that while most genes exhibit the same growth curve shape, many mutants exhibit curve shapes that deviate from the norm and fall into two categories, which we term

"stall" and "lag." Many pathways and components are enriched for deviant phenotypes, including some involved in metabolism and cellular state transitions. We also applied our method to a dynamic treatment of UV-radiation and found that deviation profiles contain inform functional gene clusters. The field of mutant screening in nutrient-restricted conditions continues to grow, and our platform – given its high throughput and information-rich measurements – is well suited for these studies.

4.6 Methods

4.6.1 Agar Preparation

Yeast Peptone Adenine Dextrose (YPAD) and Synthetic Complete (SC) agar plates were prepared following standard procedures (Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual, 2005 Edition). To maintain the *yfg*Δ::KanMX deletion, plates were supplemented with a final concentration of 100µg/ml G418 (Biopioneer).

4.6.2 Colony Preparation

Non-essential haploid strains were selected from the Yeast Knockout Collection (n = 5163; Open Biosystems) with the following genetic makeup: MATa; *his3*Δ1; *leu2*Δ0; *ura3*Δ0; m*et15*Δ0; *yfg*Δ::KanMX (Background: BY4741). Essential haploid strains were selected from the Yeast DAmP Library (n = 768; Open Biosystems) with the following genetic

makeup: MATa *his3*Δ1; *leu2*Δ0; *ura3*Δ0; *yfg*-DAmP::KanMX (Background: BY4741).

4.6.3 Plate propagation

We started with our collection in 1536 format. These were up-scaled to 6144 density and grown for about 4 hours at 30 degrees. These were then repined using the 6144-pin tool. These repined, 6144-format plates were grown for 24 hours at room temperature and then put into the cold room for two days. An additional plate containing only the HO deletion strain in 6144 format was prepared and passaged through the following experiments alongside the deletion collection.

All agar plates were poured on the same day, left untouched on the bench for 24 hours, then stacked and left on the bench for another 24 hours before being bagged and placed in the cold room.

Previous to each day of the experiment, the target plates for the following day were removed from the cold room and warmed to room temperature on the bench overnight. With the exception of the first day, for which the source plates came from the cold room, the source plates for the day's experiments had been growing on the bench overnight. These were pinned onto 8 plates: 4 SC (2 SC and 2 SC+UV), 2 SC+GAL, and 2 YPAD. The HO-deletion plate was pinned onto SC. Each plate was also pinned onto an addition SC plate to become the source plate for the

following day. The sources for the next day were bagged and placed on the bench. The condition plates were imaged.

Due to human oversight, the target plates for the second day were not warmed on the bench overnight, but were removed that morning and warmed in the 30-degree incubator for 30 minutes before pinning.

The SC+GAL data were not included in this study.

4.6.4 Experimental setup

Plates were placed facedown without lids inside the imaging light-box on a sanded, black acrylic surface in a 3x3 grid. They were imaged with a Nikon D800e camera fitted with an AF Micro Nikon 60mm lens using Nikon's Camera Control Pro 2 software. 720 images were taken at 2-minute intervals using a fixed focus (set initially using auto-focus), aperture priority with aperture at f/5 and exposure compensation at +1 2/3 EV (note: days 1 and 2 were shot using shutter priority with shutter speed at 1/10). Images were saved in NEF format (Nikon's raw image format). The guiding principle used in selecting these settings was to produce images in which the transition from agar to colony covered a larger range in the camera's saturation capability. This produces images that appear to be overly bright, but the colony size information is superior to dimmer images.

For the SC+UV condition, plates were taken from the setup immediately after image 150 was taken (5 hours), placed into a UV cross-

linker and treated with 15,000 micro-Joules (dosage was chosen over the course of multiple pilot experiments as the amount resulting in little change to most colonies, but still evoking a response from known UV-sensitive mutants). They were immediately placed back into the setup before the next image was taken.

4.6.5 Image analysis

NEF image files were converted to mosaicked TIFF using DCRAW. TIFF images were de-mosaicked using MATLAB's *demosaic* function (although options exist that avoid using the Image Analysis Toolkit for those without it). Processed images were quantified using the Matlab Colony Analyzer Toolkit [14]. The crop and colony grid placement were defined manually for each plate and were reused for each image. The RawLocalFitted algorithm was used for pixel thresholding. The ColonyArea algorithm was used for colony size quantification. Colony sizes were converted to radius, spatially corrected, Gaussian smoothed, and normalized to initial colony size. Growth curve data were reduced from 720 to 96 points using interpolation to decrease computational load for downstream analyses.

4.6.6 Quality Control

Data for some colonies were removed due to noise or artifacts that could not be corrected. All colonies on the border of the plate were

removed. Colonies in sparse regions were removed. Colonies obscured by scratches in the plastic were also removed. In the end, we retained information for about 5,000 colonies in each condition.

4.6.7 Growth-curve analysis

Normalized growth curves were averaged across the 12 replicates. Reference curves (representing "expected" or non-phenotype growth) were computed by normalizing all curves by their endpoint and by taking the mode across normalized curves for each point in time. For each curve, the difference between the normalized curve and the reference is taken. Subtle biases correlated with final colony size were corrected. The total deviation was computed as the mean sum of squared residuals. The sign was determined as the sign of the average deviation.

4.6.8 HO-deletion control and FDR

In addition to the 4 conditions already mentioned, we included 6 replicates of a plate containing 6144 copies of the HO-deletion strain as a control. We scored these data using the same procedures for the rest of the gene deletion collection, and the resulting distribution of scores served as a null distribution for total deviation scores in the SC condition.

4.6.9 GO Enrichment

Many methods exist for performing Gene Ontology[23] (GO) term enrichment – we found more practical and reliable to write our own than to find an existing method with the desired qualities. Given a list of scores for each gene, we computed the average score for each GO term, ignoring missing values. We permuted the scores and recomputed the average score for each term 1,000 times, creating a null distribution of random scores. Each score was assigned a false discovery rate (FDR) based on the null distribution of scores of the same term size. This method does not require that a score cutoff be applied.

4.7 Acknowledgments

Chapter 3, in part, is currently being prepared for submission for publication of the material. Bean, Gordon; Jaeger, Philipp; Ideker, Trey. The dissertation author was the primary investigator and author of this material.

4.8 Figures

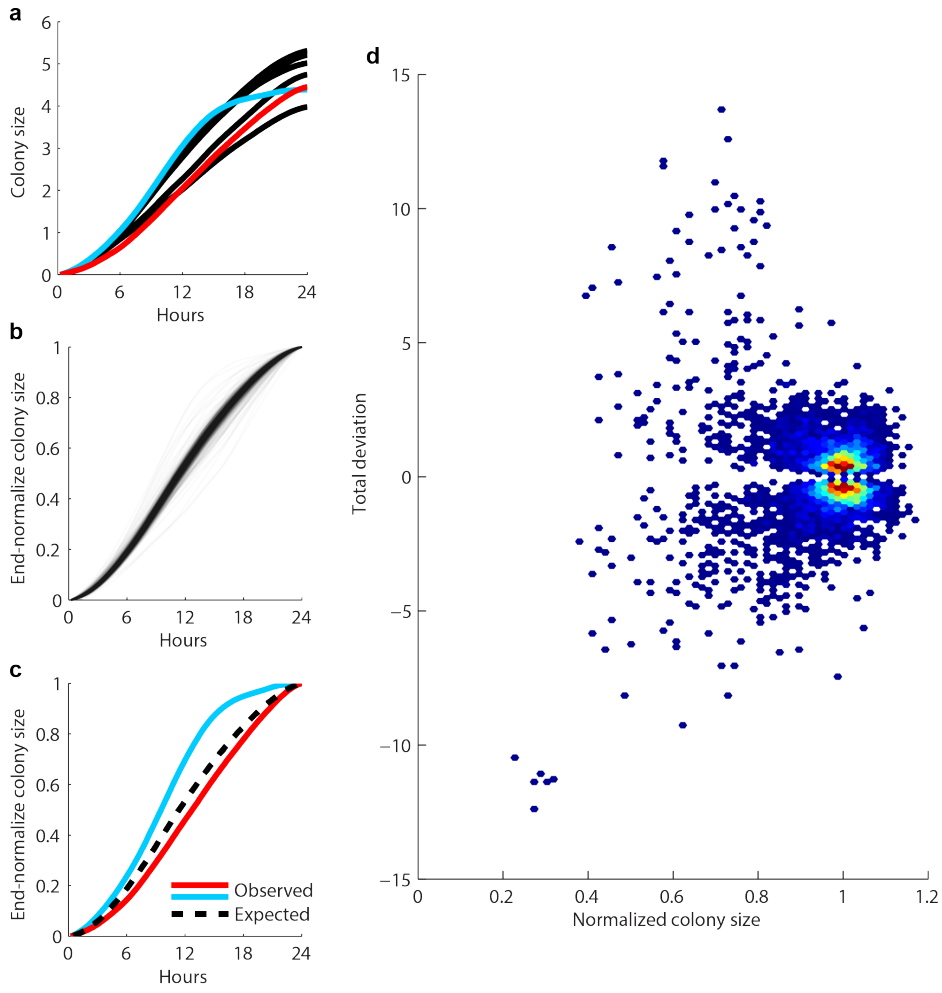


Figure 4.1 Computing the quantitative measure of deviant growth behavior.
(a) Line plots of sample growth curves; curves exhibiting deviant behavior highlighted in blue and red; colony size = length of colony radius in pixels. **(b)** Line plot of all growth curves, normalized by endpoint; opacity of each curve set to 0.01. **(c)** Line plot demonstrating how curves are compared against the reference; blue and red lines correspond to the same in (a); dashed line indicates indicate the consensus curve shape suggested in (b). **(d)** Density scatter plot of normalized colony size (y-axis; the classic measure) against total deviation (x-axis; the root-mean-squared-error of a curve to the reference); red indicates high density, blue indicates low density/single points.

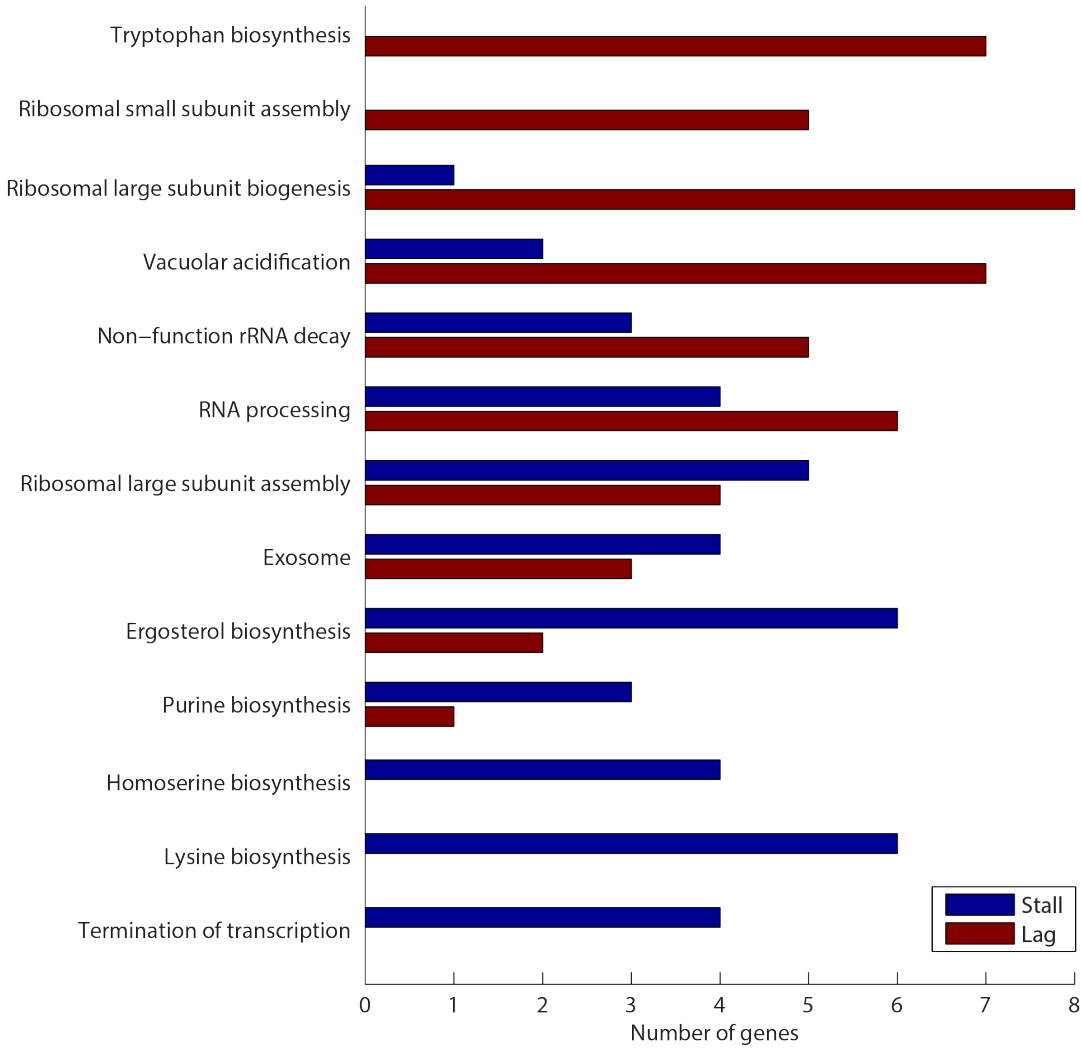


Figure 4.2 Cellular processes enriched for stall and lag phenotypes.
Bar plots of pathways or components enriched for deviant phenotypes.

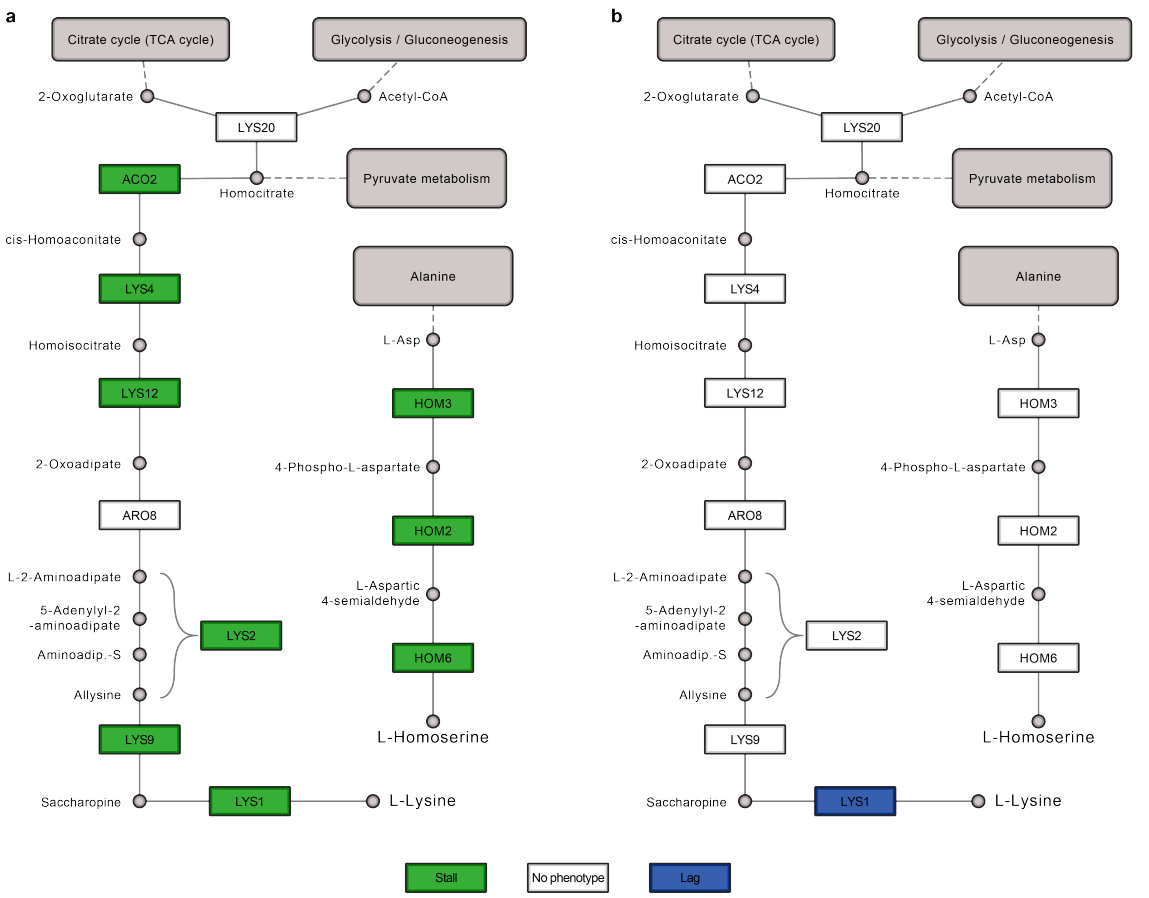


Figure 4.3 Changes in pathway phenotypes in response to different media. Diagram showing the lag and stall phenotypes mapped onto the lysine and homoserine biosynthesis pathways when grown on SC (a) and YPAD (b) media; networks adapted from KEGG Pathway sce00300.

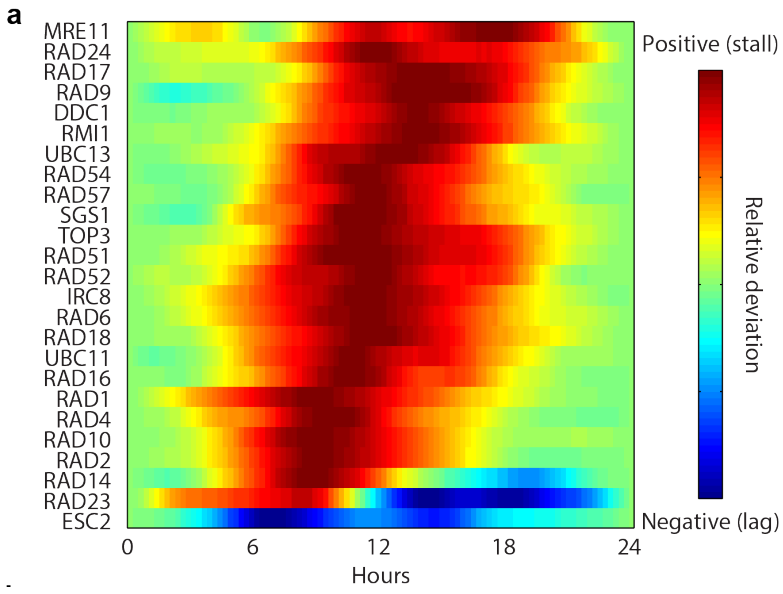


Figure 4.4 Time-lapse analysis of the UV-response.
(a) Heatmap showing the clustered deviation profiles (observed minus expected growth curves) for top UV-sensitive mutants; colorbar indicates relative magnitude of deviation values. **(b)** Heatmap showing the genetic profile similarity between for the same mutants in (a); colorbar indicates the magnitude of the profile similarity score; data from [1], with missing data shown in gray. **(c)** Bar plots of the untreated and UV-treated genetic profile similarity scores for RAD23 with RAD14 (blue) and RAD4 (red); data from [8].

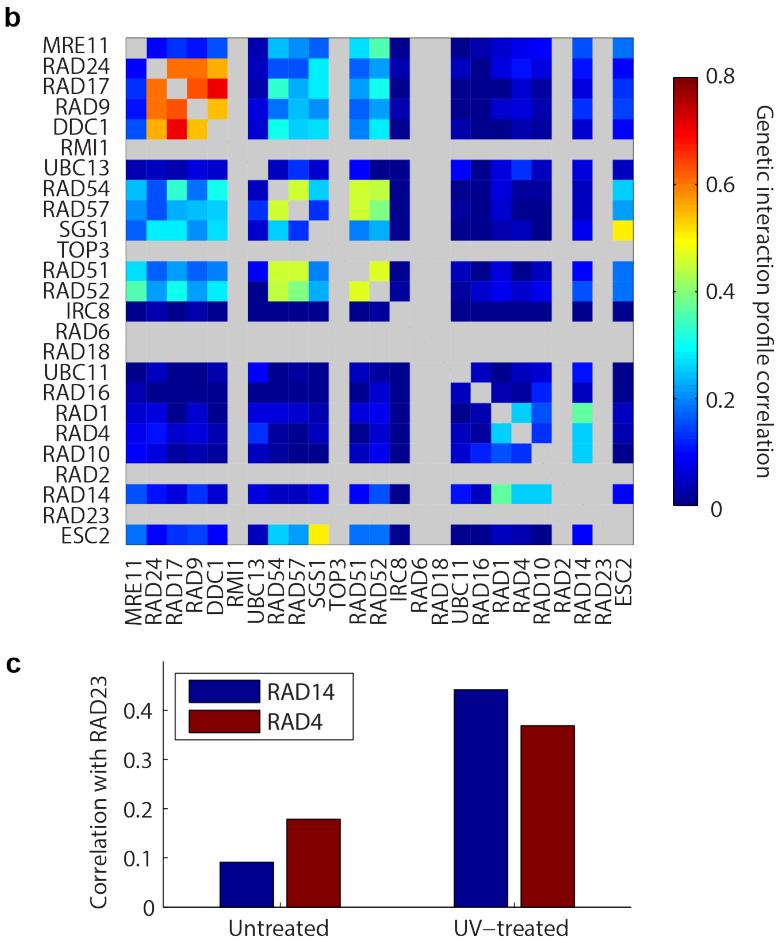


Figure 4.4 Time-lapse analysis of the UV-response. Continued.

4.9 References

1. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Tong AH, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C. The genetic landscape of a cell. *Science* 2010, 327:425–3110.1126/science.1180823.

2. Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, Weissman JS, Krogan NJ: Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 2005, 123:507–1910.1016/j.cell.2005.08.031.

3. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, Schuldiner M, Gebbia M, Recht J, Shales M, Ding H, Xu H, Han J, Ingvarsdottir K, Cheng B, Andrews B, Boone C, Berger SL, Hieter P, Zhang Z, Brown GW, Ingles CJ, Emili A, Allis CD, Toczyski DP, Weissman JS, Greenblatt JF, Krogan NJ: Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 2007, 446:806–1010.1038/nature05649.

4. Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, Silva AC, Shales M, Collins SR, van Wageningen S, Kemmeren P, Holstege FCP, Weissman JS, Keogh M-C, Koller D, Shokat KM, Krogan NJ: Functional organization of the S. cerevisiae phosphorylation network. *Cell* 2009, 136:952–6310.1016/j.cell.2008.12.039.

5. Zheng J, Benschop JJ, Shales M, Kemmeren P, Greenblatt J, Cagney G, Holstege F, Li H, Krogan NJ: Epistatic relationships reveal the functional organization of yeast transcription factors. *Mol. Syst. Biol.* 2010, 6:42010.1038/msb.2010.77.

6. Bandyopadhyay S, Mehta M, Kuo D, Sung M-K, Chuang R, Jaehnig EJ, Bodenmiller B, Licon K, Copeland W, Shales M, Fiedler D, Dutkowski J, Guénolé A, van Attikum H, Shokat KM, Kolodner RD, Huh W-K, Aebersold R, Keogh M-C, Krogan NJ, Ideker T: Rewiring of genetic networks in response to DNA damage. *Science* 2010, 330:1385–910.1126/science.1195618.

7. Guénolé A, Srivas R, Vreeken K, Wang ZZ, Wang S, Krogan NJ, Ideker T, van Attikum H: Dissection of DNA damage responses using multiconditional genetic interaction maps. *Mol. Cell* 2013, 49:346–5810.1016/j.molcel.2012.11.023.

8. Srivas R, Costelloe T, Carvunis A-R, Sarkar S, Malta E, Sun SM, Pool M, Licon K, van Welsem T, van Leeuwen F, McHugh PJ, van Attikum H, Ideker T: A UV-induced genetic network links the RSC complex to nucleotide excision repair and shows dose-dependent rewiring. *Cell Rep.* 2013, 5:1714–2410.1016/j.celrep.2013.11.035.

9. Blomberg A: Measuring growth rate in high-throughput growth phenotyping. *Curr. Opin. Biotechnol.* 2011, 22:94–10210.1016/j.copbio.2010.10.013.

10. Shah N a, Laws RJ, Wardman B, Zhao LP, Hartman JL: Accurate, precise modeling of cell proliferation kinetics from time-lapse imaging and automated image analysis of agar yeast culture arrays. *BMC Syst. Biol.* 2007, 1:310.1186/1752-0509-1-3.

11. Banks A P, Lawless C, Lydall D A: A quantitative fitness analysis workflow. *J. Vis. Exp.* 2012, :1–710.3791/4018.

12. Vandersluis B, Hess DC, Pesyna C, Krumholz EW, Syed T, Szappanos B, Nislow C, Papp B, Troyanskaya OG, Myers CL, Caudy A a: Broad metabolic sensitivity profiling of a prototrophic yeast deletion collection. *Genome Biol.* 2014, 15:R6410.1186/gb-2014-15-4-r64.

13. Warringer J, Anevski D, Liu B, Blomberg A: Chemogenetic fingerprinting by analysis of cellular growth dynamics. *BMC Chem. Biol.* 2008, 8:310.1186/1472-6769-8-3.

14. Bean GJ, Jaeger P a, Bahr S, Ideker T: Development of ultra-high-density screening

tools for microbial "omics". *PLoS One* 2014, 9:e8517710.1371/journal.pone.0085177.

15. Collins SR, Schuldiner M, Krogan NJ, Weissman JS: A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* 2006, 7:R6310.1186/gb-2006-7-7-r63.

16. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, Youn J-Y, Ou J, San Luis B-J, Bandyopadhyay S, Hibbs M, Hess D, Gingras A-C, Bader GD, Troyanskaya OG, Brown GW, Andrews B, Boone C, Myers CL: Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* 2010, 7:1017–2410.1038/nmeth.1534.

17. Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A: High-resolution yeast phenomics resolves different physiological features in the saline response. *Proc. Natl. Acad. Sci. U. S. A.* 2003, 100:15724–910.1073/pnas.2435976100.

18. Guzder SN, Sung P, Prakash L, Prakash S: Affinity of Yeast Nucleotide Excision Repair Factor 2, Consisting of the Rad4 and Rad23 Proteins, for Ultraviolet Damaged DNA. *J. Biol. Chem.* 1998, 273:31541–3154610.1074/jbc.273.47.31541.

19. Jansen L, Verhage RA, Brouwer J: Preferential Binding of Yeast Rad4 Rad23 Complex to Damaged DNA. *J. Biol. Chem.* 1998, 273:33111–3311410.1074/jbc.273.50.33111.

20. Rodriguez K, Talamantez J, Huang W, Reed SH, Wang Z, Chen L, Feaver WJ, Friedberg EC, Tomkinson AE: Affinity purification and partial characterization of a yeast multiprotein complex for nucleotide excision repair using histidine-tagged Rad14 protein. *J. Biol. Chem.* 1998, 273:34180–9.

21. Guzder SN, Bailly V, Sung P, Prakash L, Prakash S: Yeast DNA repair protein RAD23 promotes complex formation between transcription factor TFIIH and DNA damage recognition factor RAD14. *J. Biol. Chem.* 1995, 270:8385–8.

22. Cooper SJ, Finney GL, Brown SL, Nelson SK, Hesselberth J, MacCoss MJ, Fields S: High-throughput profiling of amino acids in strains of the Saccharomyces cerevisiae deletion collection. *Genome Res.* 2010, 20:1288–9610.1101/gr.105825.110.

23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 2000, 25:25–910.1038/75556.

CHAPTER 5

CONCLUSION

Three patterns emerge from the development of high-throughput genetic interaction screening in yeast. First is the tango between experimental and computational advances – an experimental improvement generates new computational problems, and solutions to these problems opens the door for new experimental improvements. First came the SGA mating technology [1], then the EMAP analysis and scoring algorithms (solving spatial artifacts, estimating expected double mutant phenotypes, and summarizing quantitative genetic interactions) [2], which in turn inspired the genome-scale interaction map [3], which necessitated new computational solutions (solving batch effects, new spatial artifacts, and neighbor effects). When differential genetic interaction mapping came about [4], more computational needs presented themselves, such as how to reliably score differential interactions and conduct further analyses on differential interaction data. In Chapter 2, I presented solutions to these challenges, introducing the dS-score and describing how differential profile similarity can identify relevant biological hypotheses.

The second pattern is the steady increase in experimental throughput. Quickly the number of colonies on a single agar plate has increased from 384 to 768 to 1536 to 6144. As the number of colonies increased, the time each colony could grow before exhausting real estate decreased. Shorter growth spans raised concerns about data quality – was there sufficient time to bring out mutant phenotypes and escape the noise introduced in the pinning process. While the 6144-colony format was about to become commercially available, there were still concerns about its utility. In Chapter 3, I and colleagues answered these concerns, showing that the 6144 format could produce reliable data and opened the door for new experimental decisions that could not be made before, such as sacrificing data quality for resource and time savings. Furthermore, I developed an open-source, customizable toolkit to address the image analysis challenges associated with the 6144-colony format. I designed the toolkit in such a way as to make it adaptable to future experimental protocols, rather than to be specifically tailored to our pipeline.

The third pattern we see is that old tools are repurposed to make new ones. The few-by-all approach used by Tong et al. [1,5] was repurposed to a many-by-many EMAP approach with distinct advantages and disadvantages [6]. The EMAP approach was later repurposed to

perform differential EMAP experiments. In these steps, the investigators identified small changes they could make to existing protocols to extract new information. In Chapter 3, I presented work that introduces the newest repurposing of the SGA/EMAP technology: time-lapse imaging and growth curve analysis. I show that through a simple modification to the standard protocol, full growth curve information can be measured. I provide a computational framework for analyzing growth curve data and quantifying new growth phenotypes.

There is also much that can be done to improvement upon my work. Differential genetic interactions are still not completely understood. In their most simple explanation, they represent only the presence and absence of interactions across conditions; however, there is evidence to suggest that changes in magnitude and changes in sign (e.g. a positive interaction becoming negative) may hold more nuanced information. Further work may be able to more clearly establish the interpretation of these interactions.

There are experimental improvements that still need to be done to optimize the 6144-colony format. Currently, border colonies – which are known to grow larger than colonies in the center of the plate, due to increased access to nutrients – are more susceptible to smearing when pinned with the 6144 pinning pad, resulting in colony contamination. Also,

the decreased depth of the 6144 pad makes the format more sensitive to agar surface topology – locations where the agar exhibits subtle depressions may result in missing colonies. Studies are under way to address these challenges.

Finally, the computational and statistical analyses of growth curves can no doubt be improved. In Chapter 4 I present a simple approach that quantifies only the degree of curve deviation and yields sound biological information. However, more specific features, such as the length of the lag phase or the timing of stall effects, could be quantified and may yield additional insights [7]. Also, the cause and interpretation of stall and lag effects can be further elucidated. Furthermore, while the mechanism for expected double-mutant phenotypes is well understood for growth-rate data, it is not clear that the same patterns will hold for growth curve data, and novel interaction models will be needed when growth-curve analysis is applied in double-mutant contexts.

My contributions have prepared the way for current and future work. The work presented in Chapters 2 and 3 have informed multiple ongoing differential epistasis mapping studies, several screens using the 6144-format are now planned or underway, and my image analysis toolkit currently provides computational solutions for several projects.

As discussed in Chapter 4, a recent study showed that examining mutant growth on special combinations of carbon and nitrogen sources provides useful metabolic information [8]. Another recent study identified interesting relationships between amino acid concentrations and various cellular components and pathways [9], suggesting a rich field of genetic and metabolic interactions to be explored. Together, media conditions and genetic backgrounds constitute hundreds of thousands to millions of unique genetic/metabolic combinations to be screened. My time-lapse approach is capable of identifying metabolic and stress-related features, and in combination with the 6144-colony format, exploring these combinations of gene background and growth environment is now feasible.

5.1 References

1. Tong a H, Evangelista M, Parsons a B, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C: Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 2001, 294:2364–810.1126/science.1065810.

2. Collins SR, Schuldiner M, Krogan NJ, Weissman JS: A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* 2006, 7:R6310.1186/gb-2006-7-7-r63.

3. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Tong AH, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C. The genetic landscape of a cell. *Science* 2010, 327:425–3110.1126/science.1180823.

4. Bandyopadhyay S, Mehta M, Kuo D, Sung M-K, Chuang R, Jaehnig EJ, Bodenmiller B, Licon K, Copeland W, Shales M, Fiedler D, Dutkowski J, Guénolé A, van Attikum H, Shokat KM, Kolodner RD, Huh W-K, Aebersold R, Keogh M-C, Krogan NJ, Ideker T: Rewiring of genetic networks in response to DNA damage. *Science* 2010, 330:1385–910.1126/science.1195618.

5. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Ménard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C.: Global mapping of the yeast genetic interaction network. *Science* 2004, 303:808–1310.1126/science.1091317.

6. Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, Weissman JS, Krogan NJ: Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 2005, 123:507–1910.1016/j.cell.2005.08.031.

7. Warringer J, Anevski D, Liu B, Blomberg A: Chemogenetic fingerprinting by analysis of cellular growth dynamics. *BMC Chem. Biol.* 2008, 8:310.1186/1472-6769-8-3.

8. Vandersluis B, Hess DC, Pesyna C, Krumholz EW, Syed T, Szappanos B, Nislow C, Papp B, Troyanskaya OG, Myers CL, Caudy A a: Broad metabolic sensitivity profiling of a prototrophic yeast deletion collection. *Genome Biol.* 2014, 15:R6410.1186/gb-2014-15-4-r64.

9. Cooper SJ, Finney GL, Brown SL, Nelson SK, Hesselberth J, MacCoss MJ, Fields S: High-throughput profiling of amino acids in strains of the Saccharomyces cerevisiae deletion collection. *Genome Res.* 2010, 20:1288–9610.1101/gr.105825.110.