

Unlocking the hidden dance of cellular resilience

Exploring the evolutionary adaptation of the cell polarity machinery in *S.cerevisiae*

Iñigo de la Cruz, L.M.

DOI

[10.4233/uuid:4f87990a-4710-4032-bc49-6b5ae4ef82e7](https://doi.org/10.4233/uuid:4f87990a-4710-4032-bc49-6b5ae4ef82e7)

Publication date

2025

Document Version

Final published version

Citation (APA)

Iñigo de la Cruz, L. M. (2025). *Unlocking the hidden dance of cellular resilience: Exploring the evolutionary adaptation of the cell polarity machinery in S.cerevisiae*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:4f87990a-4710-4032-bc49-6b5ae4ef82e7>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Unlocking the hidden dance of cellular resilience

Exploring the evolutionary adaptation of the cell polarity machinery in
S.cerevisiae

Unlocking the hidden dance of cellular resilience

Exploring the evolutionary adaptation of the cell polarity machinery in
S.cerevisiae

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. T.H.J.J. van der Hagen
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op woensdag 5 februari 2025 om 10.00 uur

door

Leila Margarita IÑIGO DE LA CRUZ

Master of Science in Applied Physics,
Technische Universiteit Delft, The Netherlands,
geboren in La Habana, Cuba.

Dit proefschrift is goedgekeurd door de

promotor: Dr.ir. L. Laan

promotor: Prof. Dr. A. M. Dogterom

Samenstelling promotiecommissie:

Rector Magnificus,

Dr.ir. L. Laan

Prof. Dr. A. M. Dogterom

voorzitter

Technische Universiteit Delft, promotor

Technische Universiteit Delft, promotor

Onafhankelijke leden:

Prof. dr. ir. S. J. Tans

Prof. dr. M. Boxem

Dr. J. van Leeuwen

Dr. J.W. Zwanikken

Prof. dr. C. Joo

Technische Universiteit Delft

Universiteit Utrecht

UMass Chan Medical School, USA

Technische Universiteit Delft

Technische Universiteit Delft, reservelid

Overige leden:

Prof. dr. E. Frey

Ludwig-Maximilians-Universität-München, Germany



Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Keywords: Functional genomics, Cellular biophysics, Functional redundancy, Protein copy number, SATAY, Genetic interactions, Epistasis, Experimental evolution, *S.cerevisiae*

Printed by: Ipskamp Printing

Designed by: Jose Carlos Urrea Llanusa

Front & Back: Artistic representation of a snapshot of cellular choreography following the music of evolution.

Copyright © 2025 by Leila Iñigo de la Cruz

ISBN/EAN: 978-94-6518-018-2

An electronic version of this dissertation is available at

<http://repository.tudelft.nl/>.

A mis mas grandes amores : Jose, Pablo y Mina

Contents

Preface	1
Summary	3
Samenvatting	5
Resumen	7
1 Introduction	9
2 Redundancy and the role of protein copy numbers in the cell polarization machinery of budding yeast	25
3 An approach to extract genetic interactions from saturated transposition analysis in yeast.	57
4 Using genetic interactions from saturated transposition analysis to uncover <i>NRP1</i>'s function in <i>S.cerevisiae</i>.	95
5 Probing the genetic rewiring of a reproducible evolutionary trajectory	131
6 Conclusions	161
Acknowledgements	169
List of Publications	173
Curriculum Vitæ	175

Preface

Once upon a time, there was a grand puzzle, but instead of traditional pieces, it was made up of tiny building blocks known as genes, proteins, and other important parts. This puzzle, much like a game of Jenga, required a precise arrangement of its pieces to maintain stability. These building blocks worked together in complex ways to ensure everything in the puzzle, or rather, in living organisms, ran smoothly. Now, understanding this puzzle was no easy feat. It was incredibly complex, and sometimes it felt like trying to solve a riddle. But scientists, with their keen eyes and sharp minds, dedicated themselves to unraveling its mysteries. What they found was truly fascinating. They discovered that even though the puzzle was intricate, it followed certain rules that helped predict how it might change over time. Just like mixing different ingredients in a recipe could create something surprisingly new and delicious, interactions between genes and proteins, known as "epistasis," led to unexpected outcomes, shaping the course of evolution. In our budding yeast Jenga game played for six years we wanted to understand how it rebounded after losing important parts, much like completing a puzzle with a few missing pieces. Through mathematical models and experiments, we uncovered the hidden mechanisms that yeast used to stay resilient, much like having spare parts for a broken car. But yeast wasn't just resilient; it was also adaptable. It could change not only its own parts but also how those parts were connected to each other. We discovered that there was redundancy in the connections between pieces, allowing the puzzle to reconfigure itself cleverly after losing a crucial piece, much like a football team adjusting its strategy when a player gets injured. Furthermore, we developed a method to predict the impact on the puzzle's stability after removing two or three pieces. Using this method, we delved deeper into the roles of specific pieces in maintaining stability and observed how removing certain pieces affected the puzzle differently. We even found that when the puzzle was almost broken and we tried to fix it by removing another piece, the consequences rippled across its pieces, with some having a higher chance of disrupting the puzzle than others. In the end, our journey through the puzzle of budding yeast revealed not only its resilience and adaptability but also the intricate dance of its pieces and the hidden rules guiding its evolution. And though our adventure may have ended, the puzzle of life continues to unfold, waiting for curious minds to continue the game.

Summary

Biological systems are dynamic and multi-layered, characterized by internal structures with varying levels of complexity that are in constant interplay. For instance, the regulated interaction between gene expression machinery and the biochemical reactions among proteins is crucial for orchestrating cellular functions. This interplay is essential for maintaining life, even in seemingly "simpler" single-cell organisms, amidst an ever-changing external environment. These interactions create a complex web of connections between different biological organization levels, making studying such systems enormously challenging. However, the story does not end there; biological systems have the remarkable ability to evolve. Evolution is fundamental to all living beings on Earth, enabling the vast diversity of forms, shapes, lifestyles, and colors observed in nature. Anticipating evolutionary outcomes has long perplexed scientists. In some cases, evolution appears to follow reproducible trajectories, providing opportunities to investigate factors that may constrain evolutionary paths while controlling for external environmental influences. One such factor, discovered in the past century, is epistasis, which refers to the variable effect of a gene mutation depending on the presence or absence of mutations in other genes. This concept has played a pivotal role in shaping our understanding of evolutionary processes. In this thesis, we examine the effects of epistatic mutations on a genome-wide scale within a specific evolutionary trajectory.

Our model system for investigating the role of epistatic mutations in evolution is the budding yeast *Saccharomyces cerevisiae*, explicitly focusing on the resilience of one of its crucial survival functions: cell polarization. Budding yeast undergoes asymmetric division by specifying a point on its membrane where the daughter cell will proliferate. Critical to this process is the precise establishment of the polarization axis. Cells lacking the *BEM1* gene fail to efficiently polarize, impeding their progression through subsequent stages of the cell cycle essential for survival. Remarkably, the function of cellular polarization exhibits surprising resilience even in the absence of critical components like *BEM1*. Previous studies have demonstrated that populations lacking the *BEM1* gene can rapidly recover by deleting two additional genes, *BEM3* and *NRP1*. This observation suggests that these genes are epistatic to *BEM1*, as their effects heavily depend on the absence of *BEM1*. Even more intriguing is the reproducibility of this recovery, with the same mutations recurrently observed in multiple experiments. What enables the cell polarization function to remain resilient following the deletion of a critical component? Furthermore, what are the implications of mutations that restore the impaired cell polarization function for other cellular functions and components? Addressing these questions will allow us to elucidate the underlying principles governing the impact of epistatic mutations on evolutionary constraints.

In **chapter 2**, we propose that multiple redundant mechanisms coexist within the protein network governing cell polarization, contributing to the module's resilience and adaptability. We discovered that redundancy in cell polarization does not stem from individual components or interactions but rather emerges at the functional process level. Even if one submodule becomes non-functional, the combined action of the remaining submodules maintains an operational mechanism for cell polarization. This resilience is achieved by fine-tuning the protein count within the

cell, ensuring that the remaining submodules can compensate effectively.

In **chapter 3**, we utilize the capabilities of Saturated Transposition Analysis in Yeast (SATAY) to investigate how epistatic mutations occurring during the studied evolutionary trajectory impact the cell at a genome-wide scale. We describe our methodology for deciphering the genotype-phenotype map using SATAY and demonstrate how this map can be leveraged to predict epistatic interactions among genes at a genome-wide scale.

Chapter 4 explores one of the applications of the genotype-phenotype map, focusing on discovering the potential functions of a specific gene. We investigate the role of Nrp1, a protein with an RNA binding domain (RBD) that lacks an assigned biological function but has been shown to suppress the deleterious effect of deleting *BEM1* in cell polarity. By computing the functional enrichment of its epistatic partners, we gain insights into the potential functions of *NRP1*, particularly in crucial cellular processes such as cell division and spindle morphology.

Chapter 5 uncovers that the initial epistatic mutation along the evolutionary trajectory (*BEM3* deletion), which suppresses most of the defects in cell polarity triggered by the deletion of *BEM1*, enables a reversal of the fitness effects of the predicted epistatic partners of *BEM1*. Interestingly, the epistatic partners that emerge after the deletion of *BEM3* are unrelated to cell polarity but are associated with other cellular processes, such as vesicle traffic and chromosome segregation. We observed that the epistatic partners of *BEM1* exhibit overlapping functions which shows that predicting compensatory mutations to $\Delta bem1$ mutants is a challenging task. However, we speculate that due to the reversion of fitness effects of multiple mutations following a compensatory mutation, such as *BEM3*, there is a transition to a system that exhibits a higher degree of functional modularity akin to the unperturbed cell. This transition facilitates the prediction of subsequent compensatory mutations. Our findings offer insights into the mechanisms underlying adaptability and evolution in biological systems.

In conclusion, these findings underscore the interplay of genes and proteins in shaping cellular functions, emphasizing the importance of considering genetic and epistatic interactions at a genome-wide scale in predictive models of evolutionary dynamics.

Samenvatting

Biologische systemen zijn dynamisch en meerlagig, met interne structuren die worden gekenmerkt door voortdurend veranderende niveaus van complexiteit. Zo dragen bijvoorbeeld de gen-expressiemachine, biochemische reacties tussen eiwitten en de wisselwerking van cellulaire functies, zelfs in ogenschijnlijk "eenvoudigere" eencellige organismen, bij aan het in stand houden van het leven te midden van een voortdurend veranderende externe omgeving. Deze interacties creëren een complex web van verbindingen tussen verschillende biologische organisatieniveaus, waardoor het bestuderen van dergelijke systemen enorm uitdagend is. Maar daar eindigt het verhaal niet; biologische systemen hebben het opmerkelijke vermogen om te evolueren. Evolutie is fundamenteel voor alle levende wezens op aarde, waardoor de immense diversiteit aan vormen, vormen, levensstijlen en kleuren in de natuur mogelijk is. Het anticiperen op evolutionaire uitkomsten heeft wetenschappers lange tijd verward. In sommige gevallen lijkt evolutie reproduceerbare trajecten te volgen, waardoor kansen ontstaan om factoren te onderzoeken die evolutionaire paden kunnen beperken terwijl externe omgevingsinvloeden worden gecontroleerd. Een van deze factoren, ontdekt in de afgelopen eeuw, is epistase, wat verwijst naar het variabele effect van een genmutatie afhankelijk van de aanwezigheid of afwezigheid van mutaties in andere genen. Dit concept heeft een cruciale rol gespeeld bij het vormgeven van ons begrip van evolutionaire processen. In deze scriptie onderzoeken we de effecten van epistatische mutaties op genome-schaal binnen een specifieke evolutionaire traject.

Ons model systeem voor het onderzoeken van de rol van epistatische mutaties in evolutie is de budding yeast *Saccharomyces cerevisiae*, waarbij we ons expliciet richten op de veerkracht van een van zijn cruciale overlevingsfuncties: celpolarisatie. Budding yeast ondergaat asymmetrische deling door een punt op zijn membraan aan te geven waar de dochtercel zal prolifereren. Essentieel voor dit proces is het precieze vaststellen van de polarisatie-as. Cellen die het gen BEM1 missen, polariseren niet efficiënt, waardoor hun voortgang door verdere stadia van de celcyclus die essentieel zijn voor overleving, wordt belemmerd. Opmerkelijk genoeg vertoont de functie van celpolarisatie verrassende veerkracht, zelfs in afwezigheid van kritieke componenten zoals BEM1. Eerdere studies hebben aangetoond dat populaties die het BEM1-gen missen, zich snel kunnen herstellen door twee extra genen te deleten, BEM3 en NRP1. Deze observatie suggereert dat deze genen epistatisch zijn aan BEM1, aangezien hun effecten sterk afhankelijk zijn van de afwezigheid van BEM1. Nog intrigerender is de reproduceerbaarheid van dit herstel, waarbij dezelfde mutaties herhaaldelijk worden waargenomen in meerdere experimenten. Wat maakt het mogelijk dat de functie van celpolarisatie veerkrachtig blijft na het verwijderen van een kritieke component? Bovendien, wat zijn de implicaties van mutaties die de aangetaste celpolarisatiefunctie herstellen voor andere cellulaire functies en componenten? Het beantwoorden van deze vragen stelt ons in staat om de onderliggende principes te verhelderen die de impact van epistatische mutaties op evolutionaire beperkingen beheersen.

In **hoofdstuk 2** stellen we voor dat meerdere redundante mechanismen bestaan binnen het eiwitnetwerk dat celpolarisatie reguleert, wat bijdraagt aan de veerkracht en aanpasbaarheid van de module. We ontdekten dat redundantie in celpolarisatie niet voortkomt uit individuele com-

ponenten of interacties, maar eerder op het niveau van het functionele proces ontstaat. Zelfs als een submodule niet-functioneel wordt, handhaaft de gecombineerde actie van de overgebleven submodules een operationeel mechanisme voor celpolarisatie. Deze veerkracht wordt bereikt door het fijnafstemmen van het eiwitaantal binnen de cel, waardoor de overgebleven submodules effectief kunnen compenseren.

In **hoofdstuk 3** maakten we gebruik van de mogelijkheden van Saturated Transposition Analysis in Yeast (SATAY) om te onderzoeken hoe epistatische mutaties die optreden tijdens de bestudeerde evolutionaire traject de cel op genoomschaal beïnvloeden. We beschrijven onze methodologie om de genotype-fenotype map met SATAY te ontcijferen en demonstreren hoe deze map kan worden gebruikt om epistatische interacties tussen genen op genoomschaal te voorspellen.

In **hoofdstuk 4** verkennen we een van de toepassingen van de genotype-fenotype map, waarbij we ons richten op het ontdekken van de mogelijke functies van een specifiek gen. We onderzoeken de rol van *Nrp1*, een eiwit met een RNA-bindend domein (RBD) dat geen toegewezen biologische functie heeft maar is aangetoond de schadelijke effect van het deleten van *BEM1* in celpolariteit te onderdrukken. Door de functionele verrijking van zijn epistatische partners te berekenen, krijgen we inzicht in de potentiële functies van *NRP1*, met name in cruciale cellulaire processen zoals celdeling en spindle-morfologie.

Hoofdstuk 5 onthult dat de initiële epistatische mutatie langs het evolutionaire traject (*BEM3*), die de meeste defecten in de celpolariteit onderdrukt die worden veroorzaakt door de verwijdering van *BEM1*, een omkering van de fitnessseffecten van de voorspelde epistatische partners van *BEM1*. Interessant is dat de epistatische partners die ontstaan na het verwijderen van *BEM3* geen verband houden met celpolariteit, maar geassocieerd zijn met andere cellulaire processen, zoals blaasjesverkeer en chromosoomsegregatie. We hebben waargenomen dat de epistatische partners van *BEM1* overlappende functies vertonen, wat aantoont dat het voorspellen van compenserende mutaties voor $\Delta bem1$ mutanten een uitdagende taak is. We speculeren echter dat als gevolg van de omkering van fitnessseffecten van meerdere mutaties na een compenserende mutatie, zoals *BEM3*, er een overgang is naar een systeem dat een hogere mate van functionele modulariteit vertoont, vergelijkbaar met de onverstoorde cel. Deze overgang vergemakkelijkt de voorspelling van daaropvolgende compenserende mutaties. Onze bevindingen bieden inzicht in de mechanismen die ten grondslag liggen aan het aanpassingsvermogen en de evolutie in biologische systemen.

Concluderend vergroot dit proefschrift ons begrip van het aanpassingsvermogen en de evolutie van biologische systemen, met implicaties op diverse terreinen, zoals fundamentele wetenschap, biotechnologie en geneeskunde. De bevindingen onderstrepen het samenspel van genen en eiwitten bij het vormgeven van cellulaire functies, en benadrukken het belang van het overwegen van genetische en epistatische interacties op genoombrede schaal in voorspellende modellen van evolutionaire dynamiek.

Resumen

Los sistemas biológicos son dinámicos y multicapa, con estructuras internas caracterizadas por niveles de complejidad en constante cambio. Por ejemplo, la maquinaria de expresión génica, las reacciones bioquímicas entre proteínas y la interacción de funciones celulares, incluso en organismos unicelulares aparentemente "más simples", contribuyen a mantener la vida en medio de un entorno externo en constante cambio. Estas interacciones crean una compleja red de conexiones entre diferentes niveles de organización biológica, lo que hace que el estudio de tales sistemas sea enormemente desafiante. Sin embargo, la historia no termina ahí; los sistemas biológicos tienen la notable capacidad de evolucionar. La evolución es fundamental para todos los seres vivos en la Tierra, permitiendo la vasta diversidad de formas, estructuras, estilos de vida y colores observados en la naturaleza. La anticipación de los resultados evolutivos ha desconcertado a los científicos durante mucho tiempo. En algunos casos, la evolución parece seguir trayectorias reproducibles, brindando oportunidades para investigar factores que pueden limitar los caminos evolutivos mientras se controlan las influencias ambientales externas. Uno de esos factores, descubierto en el siglo pasado, es la epistasia, que se refiere al efecto variable de una mutación génica dependiendo de la presencia o ausencia de mutaciones en otros genes. Este concepto ha desempeñado un papel crucial en la formación de nuestra comprensión de los procesos evolutivos. En esta tesis, examinamos los efectos de las mutaciones epistáticas a escala genómica dentro de una trayectoria evolutiva específica.

Nuestro sistema modelo para investigar el papel de las mutaciones epistáticas en la evolución es la levadura *Saccharomyces cerevisiae*, centrándonos explícitamente en la resiliencia de una de sus funciones de supervivencia crucial: la polarización celular. La levadura experimenta una división asimétrica al especificar un punto en su membrana donde la célula hija proliferará. Es crítico para este proceso el establecimiento preciso del eje de polarización. Las células que carecen del gen BEM1 no pueden polarizarse eficientemente, obstaculizando su progresión a través de etapas posteriores del ciclo celular esenciales para la supervivencia. Sorprendentemente, la función de la polarización celular muestra una sorprendente resiliencia incluso en ausencia de componentes críticos como BEM1. Estudios previos han demostrado que las poblaciones que carecen del gen BEM1 pueden recuperarse rápidamente al eliminar dos genes adicionales, BEM3 y NRP1. Esta observación sugiere que estos genes son epistáticos a BEM1, ya que sus efectos dependen en gran medida de la ausencia de BEM1. Aún más intrigante es la reproducibilidad de esta recuperación, con las mismas mutaciones observadas recurrentemente en múltiples experimentos. ¿Qué permite que la función de la polarización celular permanezca resistente después de la eliminación de un componente crítico? Además, ¿cuáles son las implicaciones de las mutaciones que restauran la función de polarización celular deteriorada para otras funciones y componentes celulares? Abordar estas preguntas nos permitirá elucidar los principios subyacentes que rigen el impacto de las mutaciones epistáticas en las limitaciones evolutivas.

En el **capítulo 2**, proponemos que múltiples mecanismos redundantes coexisten dentro de la red proteica que regula la polarización celular, contribuyendo a la resistencia y adaptabilidad del módulo. Descubrimos que la redundancia en la polarización celular no proviene de compo-

nentes o interacciones individuales, sino que surge en el nivel del proceso funcional. Incluso si un submódulo se vuelve no funcional, la acción combinada de los submódulos restantes mantiene un mecanismo operativo para la polarización celular. Esta resistencia se logra ajustando finalmente el conteo de proteínas dentro de la célula, asegurando que los submódulos restantes puedan compensar de manera efectiva.

En el **capítulo 3**, utilizamos las capacidades del Análisis de Transposición Saturada en Levadura (SATAY) para investigar cómo las mutaciones epistáticas que ocurren durante la trayectoria evolutiva estudiada impactan la célula a escala genómica. Describimos nuestra metodología para descifrar el mapa genotipo-fenotipo utilizando SATAY y demostramos cómo este mapa puede aprovecharse para predecir interacciones epistáticas entre genes a escala genómica.

En el **capítulo 4**, exploramos una de las aplicaciones del mapa genotipo-fenotipo, centrándonos en descubrir las posibles funciones de un gen específico. Investigamos el papel de *Nrp1*, una proteína con un dominio de unión a ARN (RBD) que carece de una función biológica asignada pero se ha demostrado que suprime el efecto perjudicial de eliminar *BEM1* en la polaridad celular. Al investigar las funciones de los genes epistáticos a *NRP1*, obtenemos información sobre las posibles funciones de *NRP1*, especialmente en procesos celulares cruciales como la división celular y la segregación de cromosomas.

En el **capítulo 5** descubrimos que la mutación epistática inicial a lo largo de la trayectoria evolutiva (*BEM3*), que suprime la mayoría de los defectos en la polaridad celular desencadenados por la eliminación de *BEM1*, permite una reversión de los efectos de en el fenotipo de los socios epistáticos predecidos por SATAY de *BEM1*. Curiosamente, los socios epistáticos que surgen después de la eliminación de *BEM3* no están relacionados con la polaridad celular, sino que están asociados con otros procesos celulares, como el tráfico de vesículas y la segregación cromosómica. Observamos que los socios epistáticos de *BEM1* exhiben funciones superpuestas, lo que hace que las mutaciones compensatorias sean difíciles de predecir. Sin embargo, especulamos que debido a la reversión de los efectos en el fenotipo de múltiples mutaciones después de una mutación compensatoria, como *BEM3*, hay una transición a un sistema modular similar a la célula no perturbada. Esta transición facilita la predicción de mutaciones compensatorias posteriores. Nuestros hallazgos ofrecen valiosos conocimientos sobre los mecanismos subyacentes que sustentan la adaptabilidad y la evolución en los sistemas biológicos.

En conclusión, esta tesis avanza en nuestra comprensión de la adaptabilidad y la evolución de los sistemas biológicos, con implicaciones en diversos campos como la ciencia básica, la biotecnología y la medicina. Los hallazgos subrayan la interacción de genes y proteínas en la configuración de funciones celulares, enfatizando la importancia de considerar las interacciones genéticas y epistáticas a escala genómica en modelos predictivos de dinámicas evolutivas.

Introduction

The first principle is that you must not fool yourself
– and you are the easiest person to fool.

Richard Feynman

Abstract

This chapter encompasses the "why," "what," and "how" of the thesis investigation. I commence with a discussion of the general motivation of this thesis, which centers on taking a first step toward dissecting the genetic underpinnings of complex traits and diseases ("why"). Next, we define the object of our study: An evolutionary reproducible trajectory that demonstrates the robustness of the biological function of cell polarity ("what"). We follow with the description of the areas that form the core of our study: 1. role of protein copy numbers on biological functions, 2. genetic interaction profiling to uncover gene function and 3. genetic interactions changes along a compensatory mutational path. In the first area, we harness the capabilities of using a galactose inducible promoter to control the gene expression of our protein of interest. For the second and third area we employ an emergent high-throughput methodology: Saturated Transposition Analysis in yeast, designed to connect gene mutations with their fitness consequences on a genome-wide scale ("how"). Our chapter culminates with a description of every chapter's content and a glossary of pertinent terms within the context of this thesis.

1.1. *The why*: Genetic underpinnings of complex traits and diseases

General motivation

Throughout history, our innate drive to explain the world has led to the creation of scientific theories that provide fundamental principles for understanding various phenomena. These theories encompass a wide range of topics, from explaining the uniformity of falling objects like apples and iron balls [1] to understanding the generation of warmth through friction [2]. Moreover, theories have been developed to comprehend and create music [3] and understand the perception of external stimuli [4]. Additionally, we have devised theories to predict and influence the behaviors of cellular, animal, and human systems [5–8]. The use of theories extends to anticipating the properties of objects and living beings, often relying on observations of their surroundings as a key input for inference. A relatable example from daily life is how individuals are sometimes judged based on the character of their friends. Those associated with a group of party-animal friends are often perceived as such, while those belonging to a circle of intellectuals are labeled bookish. Interestingly, rational beings frequently predict a person's character based on their family, friends, lifestyle, and hobbies, even without personal acquaintance. This mindset permeates scientific practice as well. For instance, in the classification mathematical method of nearest neighbors [9], a certain point is assigned to the group of its closest neighbors in Euclidean space based on their distance. In biology, we also deduce the properties of certain molecular function, for example, cell division, by analysing the response of the cellular system to variations in the concentrations of the different components that have been associated with that function [10–12]. In addition, it is possible to infer the function of specific DNA sequences (genes) by examining their associations with related DNA sequences. These associations are called genetic interactions, encompassing all potential relationships between the query gene and other DNA sequences [13, 14].

The notion of genetic interaction dates back about a century, with its origins traced to Bateson [15]. Bateson used this concept to describe instances in which the effect of a particular DNA sequence variation, also referred to as a genetic variant, is masked by a distinct variant at another genetic locus. The manifestation of the phenotype linked to one genetic variant becomes evident only in individuals possessing a specific mutation at the second locus [15]. Central to the genetic interaction concept is the comprehension of how genetic variants influence phenotypes, as graphically illustrated in panel A) of fig. 1.1.

Through more than a century of extensive investigation, a multitude of methodologies has been employed to elucidate the intricate mechanisms by which diverse attributes of an organism, dictated by its genetic constitution, collectively contribute to specific phenotypic outcomes, also known as genetic traits. It has been seen that even if two individuals carry the same mutation responsible for conditions such as cystic fibrosis or hemophilia, the severity of the disease can substantially differ due to genetic variations in their backgrounds [16]. Moreover, genome sequencing studies have brought to light instances where individuals possessing deleterious mutations remain entirely unaffected, presumably due to other, yet-undiscovered gene variants within their genomes that confer protection, as depicted in panel B) of fig. 1.1. Furthermore, many genetic traits, like height, intelligence, and susceptibility to common diseases, are influenced by multiple genes with small individual effects. Studying mutations with weak effects can shed light on the underlying genetic architecture of such traits.

On the other hand, the type of genetic interactions sheds light on the potential functional linkage among proteins. Genetic interactions are generally classified into two groups: positive

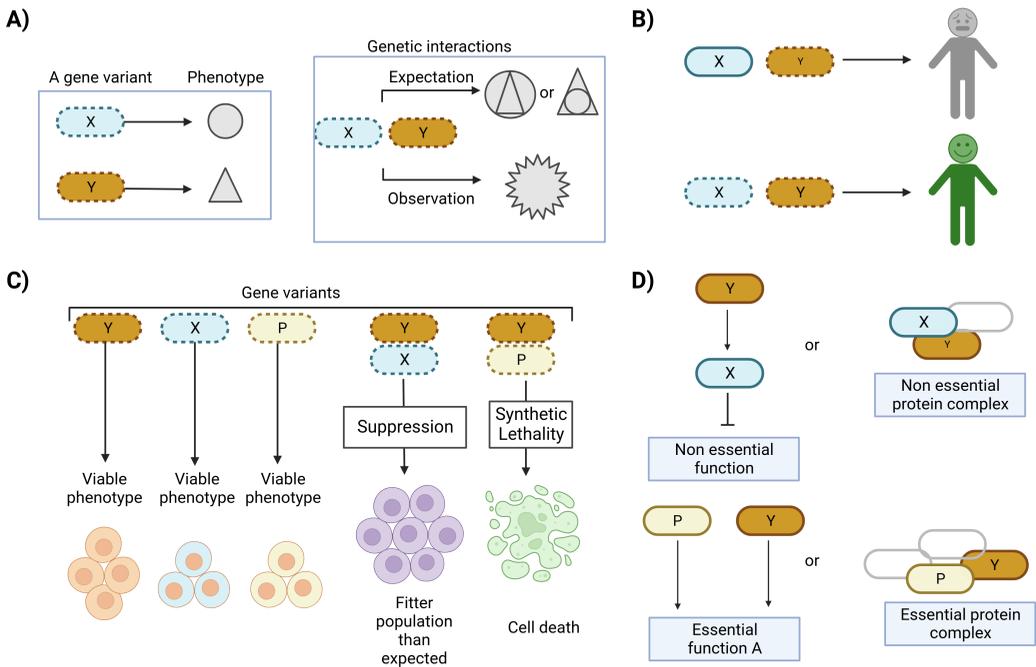


Figure 1.1. Genetic interactions. **A)** Abstract representation of the concept of genetic interactions: Oval-shaped objects symbolize genes, and the dashed stroke denotes a mutation in that gene. In the scenario where mutations in both gene X and gene Y occur within the organism, if a mutation in gene X results in a specific phenotype depicted as a circle, and a mutation in gene Y leads to a distinct phenotype represented by a square, the anticipated outcome of their combination would be the straightforward summation of their individual phenotypes. However, an unexpected combined phenotype arises when these mutations share a genetic interaction. **B)** Genetic interactions are pivotal in human health, offering insights into why certain individuals do not develop those conditions despite carrying a mutation responsible for certain diseases. The oval object lacking a dashed stroke represents a non-mutated gene. **C)** Description of extreme types of genetic interactions. When two mutations combine, and the resultant phenotype is much fitter than expected, reflected in the elevated number of cells, it is considered a suppression type of interaction. Yet, when the combination of viable mutations renders cell death, it is known as a synthetic lethality interaction. **D)** Potential functional roles of gene Y can be discerned through an analysis of its genetic interactions. Suppressive interactions observed between genes Y and X suggest their association with the same biological pathway or a non-essential protein complex. The synthetic lethality between genes Y and P suggests their involvement in compensatory pathways that contribute to an essential function or their participation within an essential protein complex.

and negative. The first concerns mutations that confer a less severe phenotype than expected, related to the organism's fitness in the presence of a second mutation. Suppression is considered an extreme type of positive interaction, where the organism fitness is at least the highest fitness individual mutation. Negative interactions refer to the opposite behavior of positive ones. Synthetic lethality is an extreme example of negative interaction in which combining two or more mutations leads to the death or non-viability of an organism or cell, even though each mutation alone would not have this effect, see panel C) from fig. 1.1.

The so-called global genetic interaction landscape is a recent effort [13, 14, 17] to dissect the effects of a particular gene loss in a broader genetic context. The landscape refers to the collec-

tion of genetic interactions between different genes within the budding yeast genome. Generally, it represents the relationships between genes that influence various aspects of cellular behavior, such as growth. The global genetic interaction landscape is represented by a network in which nodes are genes, and the edges connect genes that share a similar genetic interaction profile. It has been shown that negative interactors are usually enriched for genes that pertain to the same biological function. Moreover, negative interactors can also play redundant roles to guarantee certain essential functions or be part of the same essential complex, see panel D) from fig. 1.1. Instead, positive interactors are related to nonessential pathways or complex membership.

This tool has predicted the function of unknown genes that fall within or near functionally enriched regions and thus have a high similarity profile with the rest, for instance, the gene *IPA1*. However, for unknown genes with a poor similarity profile, it is difficult to investigate their function. Furthermore, the alteration of this network upon genetic or environmental changes is still largely unexplored [18].

1.2. *The what: On the biological system, protein copy numbers and genetic interactions*

The biological system: An evolutionary reproducible trajectory that demonstrates the robustness of the biological function of cell polarity

The cell polarity machinery in budding yeast is considered a robust functional module for several reasons, grounded in its essential role in cellular processes such as growth, division, and response to environmental cues. This machinery enables cells to establish and maintain a polarized state, which is critical for directing growth towards a specific location, ensuring proper cell division, and enabling cells to respond effectively to their environment. Furthermore, the cell polarity machinery is characterized by multiple redundant pathways and feedback loops [10, 19–22] which are crucial for its resilience, allowing the cell to adapt to various stresses and changes in the environment. In addition, the components of the cell polarity machinery are highly regulated both spatially and temporally, meaning that their activity is precisely controlled in specific regions of the cell and at specific times during the cell cycle [23–25]. This tight regulation ensures that polarity is established correctly and maintained throughout the cell cycle, contributing to the robustness of the system. Moreover, many components of the cell polarity machinery are conserved across eukaryotes [26, 27], suggesting that this system has been refined through evolution to be both effective and resilient. Lastly, on a molecular level, the proteins and pathways involved in establishing cell polarity are highly adaptable, capable of changing their behavior in response to internal and external cues [28–30]. In this thesis, we study one example of those molecular level adaptations, see fig. 1.2.

The evolutionary trajectory examined in this thesis was the primary focus of the publication: "Evolutionary adaptation after crippling cell polarization follows reproducible trajectories" by Liedewij Laan et al, from 2015. They show that budding yeast possesses evolutionary robustness against physiologically destructive perturbations. Furthermore, their findings suggest that gene deletion-driven recovery can prompt rapid divergence in the components involved in biologically significant cell functions.

This evolutionary trajectory begins with a crippled mutant bearing a full gene knockout in a nearly essential gene *BEM1*, see fig. 1.2. Yeast cells unable to express Bem1 cannot efficiently redirect the required proteins to the future bud site, resulting in an aberrant budding pattern

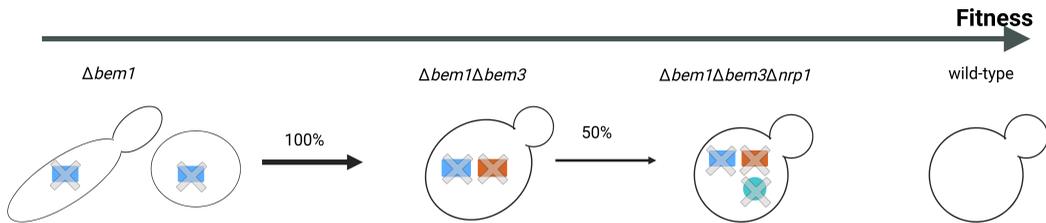


Figure 1.2. The evolutionary trajectory of yeast cells bearing a complete knockout in *BEM1* (depicted by the blue square) from [28] exhibits a consistent acquisition of a loss-of-function mutation in *BEM3* (illustrated by the orange square) across all independent experiments. This mutation significantly restores the *bem1* mutant phenotype to that resembling the wild-type. In half of these independent experiments, the gene *NRP1* (indicated by the green circle) accumulates a loss-of-function mutation, further leading to a phenotype closely resembling that of wild-type cells.

and a high death rate [28]. However, through multiple cycles of yeast culturing in a serial dilution fashion experiment [31], the population fixes three sequential mutations: *BEM3*, *NRP1*, and *BEM2*. This phenomenon is observed in nine independent lines used in the experiment. In all the independent experiments, the first mutation, *BEM3*, acquired a loss of function mutation. In five lines, the second loss of function mutation, *NRP1*, shows up. Lastly, a point mutation in the gene *BEM2* is identified only in two lines. Despite this remarkable reproducibility, sequencing studies at different generations show that eleven more mutations, composed of early stop and aminoacid substitutions, appear during the trajectory.

Furthermore, [28] shows that the order of these mutations is crucial and largely influences the population's fitness, implying that genetic interactions play a role in such evolutionary outcomes. Yet, the precise genetic and molecular mechanisms by which the first two mutations have such a degree of reproducibility and enhance the population's fitness remain puzzling.

In the aforementioned evolutionary trajectory, *Nrp1* appears as a novel protein capable of restoring the fitness of yeast cells harboring a *BEM1* knockout. Yet, a full *NRP1* knockout does not confer any prominent fitness effects in wild-type cells. Currently, its biological function and connection to polarity establishment are unknown.

We employ modeling, conceptual theory, and experiments, to the described reproducible trajectory, to manipulate the copy number of *Cdc42*, the primary regulator of budding yeast polarity. Our study suggests that multiple redundant self-organization mechanisms exist within the protein network governing cell polarization. These mechanisms contribute to the module's resilience and adaptability (chapter 2). Additionally, we use the genetic interaction profiling of *NRP1* to uncover its function since it is a gene currently associated to cell polarity but it lacks of a specific biological function. We propose that *NRP1* is linked to processes related to prion formation, endocytosis, and cell cycle checkpoints (chapter 4). Lastly, we investigate how suppression mutations alter the global genetic interaction network following an initial deleterious mutation. We analyze changes in genetic interactions along a compensatory mutational path to gain insights into this phenomenon (chapter 5).

Role of protein copy number in biological functions

Protein copy number refers to the total number of copies of a specific protein present in a cell at any given time. This measure is critical for understanding the cellular abundance of proteins,

which in turn influences biological processes, regulatory mechanisms, and the cell's overall functional state. Protein copy numbers can vary widely among different proteins within a cell, ranging from just a few copies for some regulatory proteins to several thousand copies for structural or housekeeping proteins.

Such variability in protein abundance plays a significant role in the redundancy and robustness of biological processes: high copy numbers of certain proteins can buffer the effects of mutations or loss of function in individual molecules, ensuring that the overall process or pathway remains functional [10]. Moreover, this variability also plays a role in the fine-tuning of cellular processes for instance, a high copy number of a signaling protein could amplify a signal, whereas a low copy number might make the pathway more sensitive to small changes in signal molecules [32–35].

Genetic interaction profiling to uncover gene function

One of the core aspects of this thesis is the use of genetic interaction profiles to decipher the pathways to which a particular gene may be related.

Genetic interaction profiles have been dissected for about 90% of the yeast genome by the global genetic interaction network [14, 36, 37]. This large-scale effort assigns function to specific genes by their similarity pattern of interaction profiles with other genes. Thus, it links a gene with a function following its associations with other genes through their interaction profiles. This procedure assumes genes with the same biological pathway possess strongly correlated genetic interactions.

The conventional experimental assay to determine genetic interactions has been Synthetic Genetic Arrays(SGA), which is not feasible to reproduce in standard wet labs; a more detailed explanation is in section 1.3. In this thesis, we used an alternative emergent technique to decipher genetic interactions genome-wide based on transposons called Saturated Transposition Analysis in yeast (SATAY), which is further explained in section 1.3.

Using genetic interaction profiling to understand the function of unknown genes is not a novel endeavor [13, 38, 39]. In those studies, weak effect genes appear majorly as genes with few interactions and low similarity scores with other genes. Yet, a small fraction of them are heavily connected genes at the core of essential functions like chromosome segregation and ribosome biogenesis, which may reveal those modules' high connectivity and redundancy. However, those studies are limited by the number of crossings among different single yeast knockouts; thus, genes with a poor number of interactions may be affected by this fact, which is the case of the gene *NRP1*, in which the number of crossings involving this gene spread over only 0.1% of the yeast genome, namely, about of 600 genes explored. The average number of crossings per gene in those studies represents 30% of the yeast genome.

In this thesis, I focused on the gene *NRP1* as an uncharacterized gene by the global genetic interaction network that lacks a prominent fitness effect in yeast cells to unveil its functionality through its genome-wide genetic interaction profile. Previous research [28] has demonstrated that removing this gene can restore fitness despite its neutrality in wild-type yeast cells. However, the precise mechanism underlying how the deletion of this unknown protein can restore function, in this case polarity establishment, remains elusive. Our interest lies in unraveling the function of this protein in *S.cerevisiae* and its potential linkage with the polarity module based on its genome-wide genetic interaction profile.

Genetic interactions changes along a compensatory mutational path

Genome-scale genetic interaction networks reflect the complexity of how the molecular circuitry dictates cellular behavior. Nevertheless, the degree to which a specific genetic interaction profile is modified in response to alterations in environmental or genetic conditions remains largely unexplored [18].

On the other hand, genetic interactions drive evolutionary change [40]. Thus, identifying differences in genetic interactions along a specific evolutionary mutational path can highlight how adaptive mutations influence molecular functions and cellular behavior. Furthermore, upon the dissection of how genetic interaction changes after a compensatory mutation, it is possible to generate basic principles of how evolution acts on the molecular circuitry to predict evolutionary-relevant mutations.

Current efforts in linking the global genetic network to predict evolutionary compensatory mutations have been unsuccessful [41], despite the authors claiming they used similar genetic and environmental conditions to the one used to generate the interaction network. A plausible explanation for the inability of the global genetic interaction network to successfully predict relevant compensatory mutations is lacking.

Despite genetic interaction profiles showing a more ample spectrum of possible evolutionary relevant mutations (that increase fitness), it also has the potential to anticipate forbidden possible mutational paths determined by the synthetic lethal pairs. The actual power of genetic interaction profiles to delineate principles to steer evolution remains to be seen.

In this thesis, we analyzed how the first fixed mutation along the evolutionary trajectory shapes the genetic interactions genome-wide. Moreover, we are interested in which genetic proxy serves to predict fitness and the degree of predictability that genetic interaction landscapes can have on relevant evolutionary outcomes.

1.3. *The how*: On methods to control protein copy numbers and to identify genetic interactions

Experimental procedures to control the protein copy numbers in the cell

As experimental scientists studying protein function and its impact on cellular processes, we require the ability to control and manipulate protein copy numbers within cells. Techniques such as plasmid-based overexpression or RNA interference (RNAi) can be used to increase or decrease the levels of specific proteins, respectively [42, 43]. CRISPR-Cas9 Gene Editing is another example that allows for precise genomic editing. By targeting gene loci responsible for the expression of specific proteins, researchers can create knockouts (complete removal of protein expression), knockins (increased expression by adding copies of a gene), or fine-tuned mutations that affect the gene's expression levels. However, the mentioned techniques, despite their power on manipulating gene expression, possess one important drawback which is related to off-target effects that can lead to unintended mutations, potentially affecting the cell's phenotype and viability. During our research, we decided to adopt a rather simpler technique: the application of an inducible promoter to control the expression of our gene of interest, specifically the Gal promoter [44]. The expression can be ideally tightly controlled by the addition or removal of galactose as the inducer, and thereby allows to adjust the protein's copy number in the cell dynamically.

We opted for the gal promoter to control the protein copy number of our protein of interest due to its technical simplicity compared to RNAi or CRISPR-Cas9 based techniques. The gal pro-

motor provides adjustable expression levels depending on the inducer concentration and shows minimal background expression without the inducer, reducing the risk of off-target effects or protein overexpression toxicity. However, we acknowledge a potential pitfall: leakiness, where the gene of interest is expressed even without the inducer. This could complicate experiments, particularly if the protein has a significant impact on the cell at low concentrations.

Experimental procedures to identify genetic interactions

Automated yeast genetics: Synthetic Genetic Arrays. The scientific community has dedicated significant efforts to unraveling the influence of genetic background on disease development and drug design, with essential support from extensive research on profiling the genetic interaction network in various species. A notable achievement in this pursuit is the comprehensive exploration of the global yeast interaction network by researchers such as [13, 14, 45]. The network was constructed using Synthetic Genetic Array (SGA). SGA is a high-throughput genetic screening technique used in molecular biology and genetics to analyze genetic interactions between genes in a genome systematically. It was developed to understand how genes function together to influence various cellular processes and phenotypes. The main steps of the SGA technique typically include the creation of mutant libraries, automated mating and selection of double mutants (which requires specialized robotics and automated platforms), and phenotypic analysis (an automated measure of colony size). For the global genetic interaction landscape, the authors tested 90% of the yeast genome for all possible pairwise genetic interactions, identifying nearly 1 million interactions, including around 550,000 negative and around 350,000 positive interactions [37]. SGA is a powerful and widely used technique for quantifying genetic interactions, yet it has disadvantages and limitations. First, the complexity and cost of SGA analysis are significant. Generating and analyzing numerous yeast double mutants require a sophisticated, untractable robotic infrastructure to replicate in standard wet labs. The systematic crossing of thousands of yeast mutants poses logistical challenges and incurs considerable expenses. Another limitation of SGA analysis is its inherent focus on coding sequences, potentially overlooking non-coding regions' regulatory effects on genetic interactions.

SATAY as a new emergent technique for genome-wide genetic interaction profiling. Saturated Transposition Analysis in yeast (SATAY) [39] is a technique that involves saturating the yeast genome with transposons. Transposons are mobile DNA elements discovered by Barbara McClintock in the mid-1940s. Transposons are genetic elements capable of regulating gene activity [46]. SATAY is part of the realm of methods involving transposon insertion sequencing (TIS)[47]. TIS combines genome-wide transposon mutagenesis with high-throughput sequencing to estimate the fitness contribution or essentiality of each genetic component of the species under study. Furthermore, the vast majority of TIS studies to date have been performed in bacterial species[48–50], owing to this ease of genetic manipulation. Yet, several related TIS-like methods have been developed in fungal[39, 51, 52], mammal[53–56] and archaeal[57, 58] backgrounds.

The basic concept behind SATAY is to induce the production of a transposase enzyme (Ac) that translocates the transposons (MiniDs, originally from Maize) to random locations in the genome. Usually, this system is encapsulated as a plasmid integrated into the yeast cell genome. The assay begins with constructing a saturated mutant library by introducing the randomly inserted transposon into a strain of interest, often by transformation. The goal is to create a population of yeast cells where each cell carries a single transposon insertion in the genome. Generally, when a transposon hits a genetic component, its translation gets disrupted. When cells are pooled to-

gether, each genetic component is disrupted multiple times at different sites. Probing the fitness effects of these disruptions is part of the assay, where a population containing all the possible mutations can expand for a few generations. Alternatively, the library can be subjected to a selective condition, such as gene mutation, to query non-essential features involved in survival and growth within that new genetic background. Such conditionally important components are defined by insertions whose frequency significantly changes in the population before and after the mutational change. Genomic features with disruptive transposon insertions with a decrease in frequency over experimental selection are assumed to be important for fitness in the test conditions; such features could include synthetic lethal genes with the test mutation.

Unlike Synthetic Genetic Arrays (SGA), SATAY does not rely on specific robotic setups and is not limited to coding sequences. Transposon insertions can disrupt entire genes or specific protein domains, providing valuable information about the importance of different protein regions for cellular function.

TIS has been effectively utilized to quantify the comprehensive distribution of fitness effects on a genome-wide scale and extract genetic interactions in bacterial models [59, 60]. Similarly, SATAY has proven its utility in identifying essential genes, as well as genes associated with drug resistance, such as rapamycin, and novel regulatory genes involved in the TORC1 pathway [39, 61]. These studies have also touched upon genetic interactions, although their primary focus was to demonstrate the efficacy of the SATAY technique in achieving these outcomes.

However, there exists a significant gap in research that fully exploits the potential of SATAY to systematically construct comprehensive genome-wide genetic interaction profiles, particularly in contexts such as predicting the functions of uncharacterized proteins or unraveling the genetic reorganization inherent in compensatory pathways. This thesis, therefore, represents a step towards advancing the utilization of SATAY for genome-wide mapping of genetic interactions within relevant biological systems.

1.4. Thesis aim and outline

This thesis explores the implications of the evolutionary adaptation of cell polarity from two angles: 1) Analyzing the role of copy number changes of a central regulator of cell polarity, and 2) Investigating genome-wide genetic interaction changes during the adaptation process by i. deciphering the functionality of *NRP1* as a gene with weak effects, currently uncharacterized by the global genetic interaction network, and by ii. analyzing genetic interaction changes following the main compensatory mutation of the evolutionary trajectory.

The following outline presents the goal and main results of the upcoming chapters.

Chapter 2: "Redundancy and the role of protein copy numbers in the cell polarization machinery of budding yeast."

In Chapter 2, we focus on the question: How can a self-organized cellular function evolve, adapt to perturbations, and acquire new sub-functions? We tackled this question by combining modeling, conceptual theory and experiments to propose that multiple, redundant self-organization mechanisms coexist within the protein network underlying cell polarization and are responsible for the module's resilience and adaptability.

Chapter 3: "An approach to extract genetic interactions from saturated transposition analysis in yeast."

In Chapter 3, we focus on the question: How can we implement a suitable approach to extract genetic interactions from Saturated Transposition Analysis in Yeast (SATAY)? In this approach, genetic interactions are derived using a multiplicative fitness model that accounts for fitness variances across replicates. We envision that the fitness model and resulting genetic interactions provide a foundation for guiding further investigations using low-throughput experiments.

Chapter 4: "Using genetic interactions from saturated transposition analysis to uncover *NRP1*'s function in *S.cerevisiae* ".

In Chapter 4, we focus on the question: What is the biological role of the cell polarity-associated gene *NRP1*? And by which biological mechanisms the *NRP1* knockout can compensate for the loss of *BEM1*? We tackled those questions by quantifying whether the RNA binding domain annotated to Nrp1 influences gene expression in different mutants, and by identifying genome-wide genetic interaction partners through SATAY, aiming to propose potential biological functions involving *NRP1*. We propose that *NRP1* is linked in processes related to prion formation, endocytosis, and cell cycle checkpoints, and our observations hints that *NRP1* deletion rescues $\Delta bem1$ mutants by either regulating the precise cell size at the START point of the cell cycle or enhancing the Cdc42 exchange from the membrane to the cytosol.

Chapter 5: "Probing the genetic rewiring of a reproducible evolutionary trajectory."

In chapter 5, we focus on the question: How are suppressor mutations changing the global genetic interaction map with respect to the initial deleterious mutation? Here, we utilized Saturated Transposition Analysis in Yeast (SATAY) to quantify the genetic interactions changes across genetic backgrounds. We mainly propose that genetic interactions are highly contingent on specific experimental conditions and suppressor mutation arrival predominantly induces sign epistasis changes in comparison to the initial mutant.

1.5. Glossary

DNA(Deoxyribonucleic Acid): DNA is a molecule that contains the genetic instructions used in the development, functioning, growth, and reproduction of all living organisms and many viruses. It is a long double-stranded helical structure composed of four nucleotide bases: adenine (A), thymine (T), cytosine (C), and guanine (G).

Gene: A gene is a DNA segment containing the code for a specific protein or functional RNA molecule. Genes are the fundamental units of heredity and play a crucial role in determining various traits and characteristics of an organism.

Genetic variant: A genetic variant, also known as a mutation or allele, is a specific form of a gene that differs from the prevalent form (wild-type) found in a population. Genetic variants can result from changes in nucleotide sequences, and they may or may not have an observable impact on the phenotype or traits of an individual.

Genetic locus: A genetic locus refers to the specific physical location of a gene or DNA sequence on a chromosome. It is used to identify and map the positions of genes and genetic markers in the genome.

Coding sequence: The coding sequence, also known as the coding region or open reading frame (ORF), is a specific segment of DNA or RNA that contains the instructions for producing a functional protein. In molecular genetics, this region is of particular significance as it encodes the

amino acid sequence of a protein, which, in turn, determines the protein's structure and function. The coding sequence is typically found within a gene, a segment of DNA or RNA that serves as a unit of heredity. Within a gene, the coding sequence is flanked by non-coding regions, such as promoter and terminator regions, responsible for regulating the transcription and translation of the gene.

Transposons: Transposons, also known as transposable elements or jumping genes, are segments of DNA that have the unique ability to move or transpose within a genome. They were first discovered by Barbara McClintock in the 1940s while studying the genetics of corn plants, for which she was awarded the Nobel Prize in Physiology or Medicine in 1983. Transposons make up a significant portion of many organisms' genomes, including humans. They are found in prokaryotes (bacteria) and eukaryotes (organisms with complex cell structures, including plants, animals, and fungi). These mobile genetic elements can positively and negatively impact the host organism's genome and evolution.

High-throughput techniques: High-throughput techniques refer to experimental methods and technologies that allow researchers to process large amounts of data or conduct numerous experiments rapidly and efficiently. These techniques are widely used in various scientific fields, including genomics, proteomics, drug discovery, and systems biology. The key advantage of high-throughput techniques is their ability to handle multiple samples or data points simultaneously, enabling researchers to gather comprehensive and robust data relatively quickly.

Fitness: In biology and evolutionary biology, fitness refers to the measure of an organism's ability to survive, reproduce, and pass on its genes to the next generation. It is a central concept in the theory of natural selection proposed by Charles Darwin and is fundamental to understanding evolution.

Epistasis: Epistasis refers to a phenomenon in genetics where the effect of one gene (or genetic variant) is modified by the presence or absence of alleles at another gene or genetic locus. In other words, it describes the interaction between different genes that influences the expression of a particular trait. Epistasis can manifest in various ways, including masking the effects of one gene by another, amplifying or diminishing the effects of a gene, or creating novel phenotypic outcomes that are not predictable based on the individual effects of each gene. Understanding epistasis is crucial for unraveling the complexities of genetic inheritance patterns and phenotypic diversity within populations.

Phenotype: The phenotype refers to an organism's observable physical and biochemical traits, characteristics, and behaviors. It results from interactions between an organism's genetic makeup (genotype) and environmental factors.

Evolution: Evolution is a fundamental concept in biology and refers to the gradual change and diversification of living organisms over successive generations. It is the central unifying principle in understanding the history and diversity of life on Earth.

Network: In biology or genetics, a network usually refers to a complex interconnected system involving genes, proteins, or other biological molecules. These networks can represent various interactions, such as gene regulatory networks, protein-protein interaction networks, or metabolic pathways, that collectively influence cellular functions and phenotypic outcomes. Network analysis helps researchers understand the intricate relationships and dynamics within biological systems.

Bibliography

1. Verlinde, E. On the origin of gravity and the laws of Newton. *Journal of High Energy Physics* **2011**, 1–27 (2011).
2. Quiñones-Cisneros, S. E., Schmidt, K. A., Creek, J. & Deiters, U. K. Friction theory modeling of the non-Newtonian viscosity of crude oils. *Energy & fuels* **22**, 799–804 (2008).
3. Krumhansl, C. L. Music psychology and music theory: Problems and prospects. *Music Theory Spectrum* **17**, 53–80 (1995).
4. Werner, H. & Wapner, S. Toward a general theory of perception. *Psychological review* **59**, 324 (1952).
5. Alves, N. M., Pashkuleva, I., Reis, R. L. & Mano, J. F. Controlling cell behavior through the design of polymer surfaces. *Small* **6**, 2208–2220 (2010).
6. McCaig, C. D., Rajnicek, A. M., Song, B. & Zhao, M. Controlling cell behavior electrically: current views and future potential. *Physiological reviews* (2005).
7. Chung, B. G. & Choo, J. Microfluidic gradient platforms for controlling cellular behavior. *Electrophoresis* **31**, 3014–3027 (2010).
8. Ulrich, R., Stachnik, T. & Mabry, J. *Control of human behavior* (Scott, Foresman Chicago, 1966).
9. Kramer, O. & Kramer, O. K-nearest neighbors. *Dimensionality reduction with unsupervised nearest neighbors*, 13–23 (2013).
10. Brauns, F. *et al.* Redundancy and the role of protein copy numbers in the cell polarization machinery of budding yeast. *Nature Communications* **14**, 6504 (2023).
11. De Boer, P., Crossley, R. & Rothfield, L. The essential bacterial cell-division protein FtsZ is a GTPase. *Nature* **359**, 254–256 (1992).
12. Kohyama, S., Merino-Salomón, A. & Schwillie, P. In vitro assembly, positioning and contraction of a division ring in minimal cells. *Nature Communications* **13**, 6098 (2022).
13. Boone, C., Bussey, H. & Andrews, B. J. Exploring genetic interactions and networks with yeast. *en. Nature Reviews Genetics* **8**, 437–449. ISSN: 1471-0056, 1471-0064. <http://www.nature.com/articles/nrg2085> (2021) (June 2007).
14. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**. Publisher: American Association for the Advancement of Science, aaf1420. <http://www.science.org/doi/10.1126/science.aaf1420> (2022) (Sept. 2016).
15. Bateson, W. & Mendel, G. *Mendel's principles of heredity* (Courier Corporation, 2013).
16. Hou, J., Tan, G., Fink, G. R., Andrews, B. J. & Boone, C. Complex modifier landscape underlying genetic background effects. *Proceedings of the National Academy of Sciences* **116**, 5045–5054 (2019).
17. Boone, C., Bussey, H. & Andrews, B. J. Exploring genetic interactions and networks with yeast. *Nature Reviews Genetics* **8**, 437–449 (2007).

18. Bajić, D., Moreno-Fenoll, C. & Poyatos, J. F. Rewiring of genetic networks in response to modification of genetic background. *Genome Biology and Evolution* **6**, 3267–3280 (2014).
19. Yang, S., Cope, M. J. T. & Drubin, D. G. Sla2p is associated with the yeast cortical actin cytoskeleton via redundant localization signals. *Molecular biology of the cell* **10**, 2265–2283 (1999).
20. Li, R. & Murray, A. W. Feedback control of mitosis in budding yeast. *Cell* **66**, 519–531 (1991).
21. Tang, Q. *et al.* WAVE1 and WAVE2 have distinct and overlapping roles in controlling actin assembly at the leading edge. *Molecular biology of the cell* **31**, 2168–2178 (2020).
22. Daalman, W. K.-G., Sweep, E. & Laan, L. The Path towards Predicting Evolution as Illustrated in Yeast Cell Polarity. *Cells* **9**, 2534 (2020).
23. Tschirpke, S., Daalman, W. K. & Laan, L. Quantification of GTPase cycling rates of GTPases and GTPase: effector mixtures using GTPase Glo™ assays. *bioRxiv*, 2023–11 (2023).
24. Tschirpke, S., van Opstal, F., van der Valk, R., Daalman, W. K. & Laan, L. A guide to the in vitro reconstitution of Cdc42 GTPase activity and its regulation. *bioRxiv*, 2023–04 (2023).
25. Etienne-Manneville, S. Cdc42-the centre of polarity. *Journal of cell science* **117**, 1291–1300 (2004).
26. Diepeveen, E. T., Gehrmann, T., Pourquié, V., Abeel, T. & Laan, L. Patterns of conservation and diversification in the fungal polarization network. *Genome Biology and Evolution* **10**, 1765–1782 (2018).
27. Diepeveen, E. T., de la Cruz, L. I. & Laan, L. Evolutionary dynamics in the fungal polarization network, a mechanistic perspective. *Biophysical Reviews* **9**, 375–387 (2017).
28. Laan, L., Koschwanez, J. H. & Murray, A. W. Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *Elife* **4**, e09638 (2015).
29. Nelson, W. J. Adaptation of core mechanisms to generate cell polarity. *Nature* **422**, 766–774 (2003).
30. Chiou, J.-g., Balasubramanian, M. K. & Lew, D. J. Cell polarity in yeast. *Annual review of cell and developmental biology* **33**, 77–101 (2017).
31. Popa-Burke, I. *et al.* Establishing quality assurance criteria for serial dilution operations on liquid-handling equipment. *SLAS Discovery* **14**, 1017–1030 (2009).
32. Njenga, R., Boele, J., Öztürk, Y. & Koch, H.-G. Coping with stress: how bacteria fine-tune protein synthesis and protein transport. *Journal of Biological Chemistry*, 105163 (2023).
33. Quax, T. E., Claassens, N. J., Söll, D. & van der Oost, J. Codon bias as a means to fine-tune gene expression. *Molecular cell* **59**, 149–161 (2015).
34. Shahrezaei, V. & Marguerat, S. Connecting growth with gene expression: of noise and numbers. *Current opinion in microbiology* **25**, 127–135 (2015).
35. Liu, L. *et al.* Mitophagy receptor FUNDC1 is regulated by PGC-1 α /NRF1 to fine tune mitochondrial homeostasis. *EMBO reports* **22**, e50629 (2021).
36. Kuzmin, E. *et al.* -SGA: synthetic genetic array analysis for systematically screening and quantifying trigenic interactions in yeast. en. *Nature Protocols* **16**, 1219–1250. ISSN: 1754-2189, 1750-2799. <http://www.nature.com/articles/s41596-020-00456-3> (2021) (Feb. 2021).

37. Usaj, M. *et al.* TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. *G3 Genes | Genomes | Genetics* **7**, 1539–1549. ISSN: 2160-1836. <https://doi.org/10.1534/g3.117.040220> (2022) (May 2017).
38. Van Leeuwen, J., Boone, C. & Andrews, B. J. Mapping a diversity of genetic interactions in yeast. *Current opinion in systems biology* **6**, 14–21. ISSN: 2452-3100. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6269142/> (2022) (Dec. 2017).
39. Michel, A. H. *et al.* Functional mapping of yeast genomes by saturated transposition. *en. eLife* **6**, e23570. ISSN: 2050-084X. <https://elifesciences.org/articles/23570> (2021) (May 2017).
40. Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535–538 (2012).
41. Klim, J., Zielenkiewicz, U., Kurlandzka, A. & Kaczanowski, S. The Adaptive Landscape of Genetic Interaction Network Has No Impact on Yeast Adaptive Evolution. *Frontiers in Genetics* **12**, 640501 (2021).
42. Drinnenberg, I. A. *et al.* RNAi in budding yeast. *Science* **326**, 544–550 (2009).
43. Nakayashiki, H. & Nguyen, Q. B. RNA interference: roles in fungal biology. *Current opinion in microbiology* **11**, 494–502 (2008).
44. Hovland, P., Flick, J., Johnston, M. & Sclafani, R. A. Galactose as a gratuitous inducer of GAL gene expression in yeasts growing on glucose. *Gene* **83**, 57–64 (1989).
45. Kuzmin, E. *et al.* τ -SGA: synthetic genetic array analysis for systematically screening and quantifying trigenic interactions in yeast. *Nature protocols* **16**, 1219–1250 (2021).
46. McClintock, B. Some parallels between gene control systems in maize and in bacteria. *The American Naturalist* **95**, 265–277 (1961).
47. Cain, A. K. *et al.* A decade of advances in transposon-insertion sequencing. *en. Nature Reviews Genetics* **21**. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 9 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Bacterial genetics;High-throughput screening;Mobile elements;Mutagenesis Subject_term_id: bacterial-genetics;high-throughput-screening;mobile-elements;mutagenesis, 526–540. ISSN: 1471-0064. <https://www.nature.com/articles/s41576-020-0244-x> (2021) (Sept. 2020).
48. Christen, B. *et al.* The essential genome of a bacterium. *Molecular systems biology* **7**, 528 (2011).
49. Girgis, H. S., Liu, Y., Ryu, W. S. & Tavazoie, S. A comprehensive genetic characterization of bacterial motility. *PLoS genetics* **3**, e154 (2007).
50. Van Opijnen, T., Bodi, K. L. & Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature methods* **6**, 767–772 (2009).
51. Guo, Y. *et al.* Integration profiling of gene function with dense maps of transposon integration. *Genetics* **195**, 599–609 (2013).
52. Lee, S. Y. *et al.* Dense transposon integration reveals essential cleavage and polyadenylation factors promote heterochromatin formation. *Cell reports* **30**, 2686–2698 (2020).
53. Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231–1235 (2009).

54. Carette, J. E. *et al.* Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nature biotechnology* **29**, 542–546 (2011).
55. Friedrich, M. J. *et al.* Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. *Nature protocols* **12**, 289–309 (2017).
56. Rad, R. *et al.* A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. *Nature genetics* **47**, 47–56 (2015).
57. Zhang, C., Phillips, A., Wipfler, R., Olsen, G. & Whitaker, R. *The essential genome of the crenarchaeal model Sulfolobus islandicus*. *Nat Commun* **9**: 4908 2018.
58. Sarmiento, F., Mrázek, J. & Whitman, W. B. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proceedings of the National Academy of Sciences* **110**, 4726–4731 (2013).
59. DeJesus, M. A. *et al.* Statistical analysis of genetic interactions in Tn-Seq data. en. *Nucleic Acids Research* **45**, e93–e93. ISSN: 0305-1048, 1362-4962. <https://academic.oup.com/nar/article/45/11/e93/3044354> (2021) (June 2017).
60. DeJesus, M. A. & Ioerger, T. R. Normalization of transposon-mutant library sequencing datasets to improve identification of conditionally essential genes. en. *Journal of Bioinformatics and Computational Biology* **14**, 1642004. ISSN: 0219-7200, 1757-6334. <https://www.worldscientific.com/doi/abs/10.1142/S021972001642004X> (2021) (June 2016).
61. Chen, P., Michel, A. H. & Zhang, J. Transposon insertional mutagenesis of diverse yeast strains suggests coordinated gene essentiality polymorphisms. *Nature Communications* **13**, 1490 (2022).

Redundancy and the role of protein copy numbers in the cell polarization machinery of budding yeast

Strength comes from waiting.

Jose Marti

Abstract How can a self-organized cellular function evolve, adapt to perturbations, and acquire new sub-functions? To make progress in answering these basic questions of evolutionary cell biology, we analyze, as a concrete example, the cell polarity machinery of *Saccharomyces cerevisiae*. This cellular module exhibits an intriguing resilience: it remains operational under a genetic perturbation and recovers quickly and reproducibly from the deletion of one of its key components. Using a combination of modeling, conceptual theory, and experiments, we propose that multiple, redundant self-organization mechanisms coexist within the protein network underlying cell polarization and are responsible for the module's resilience and adaptability. Based on our mechanistic understanding of polarity establishment, we hypothesize how scaffold proteins, by introducing new connections in the existing network, can increase the redundancy of mechanisms and thus increase the evolvability of other network components. Moreover, our work gives a perspective on how a complex, redundant cellular module might have evolved from a more rudimentary ancestral form.

Chapter published as a peer-reviewed publication: Brauns, F., Iñigo de la Cruz, L., Daalman, W.K.G. et al. Redundancy and the role of protein copy numbers in the cell polarization machinery of budding yeast. *Nat Commun* 14, 6504 (2023).

Author Contributions: FB, LIC, WD, JH, LL, and EF designed research; FB, JH, and EF designed the theoretical models and performed the mathematical analyses; LIC, WD, and IB and LL designed and carried out the experiments; FB, LIC, WD, LL and EF wrote the paper.

2.1. Introduction

Biological systems are self-organized. Their function emerges by the collective interplay of many components—governed by physical and chemical processes. How do such collective (self-organized) functions evolve and adapt to strong perturbations such as the loss of essential components [1, 2]?

A striking example for such adaptation is the Cdc42 cell-polarization machinery of *Saccharomyces cerevisiae* (budding yeast). Cell polarization directs cell division of budding yeast through the formation of a polar zone with high Cdc42 concentration on the membrane (see fig. 2.1 a-c). Following the knock-out of Bem1, a key player in the Cdc42-interaction network (fig. 2.1 d.), cells regain their ability to polarize and divide by loss of another component of this network. This happens rapidly (within 100 generations) and reproducibly [3]. How this recovery works has remained unclear.

Cell polarization of budding yeast is organized by a complex interaction network (fig. 2.1 d.) around the central polarity protein Cdc42. Cdc42 is a GTPase that cycles between an active (GTP-bound) and an inactive (GDP-bound) state. The key features of these two states are that active Cdc42 is strongly membrane bound and recruits many downstream factors, while inactive Cdc42-GDP can detach from the membrane to the cytosol where it diffuses freely.

In wild-type (WT) cells, polarization is directed by upstream cues like the former bud-scar [9–12]. Importantly however, Cdc42 can polarize *spontaneously* in a random direction in the absence of such cues [13–15]. What are the elementary processes underlying spontaneous Cdc42 polarization? On the timescale of polarity establishment, the total protein copy number of Cdc42 proteins (as well as its interaction partners) is nearly constant. Hence, to establish a spatial pattern in the protein concentration, the so-called polar zone, the proteins need to be spatially redistributed in the cell by *directed transport*. There are two distinct, mostly independent, pathways for directed transport that have been established by experimental and theoretical studies [4, 14–17]: cytosolic diffusive flux driven by a sustained concentration gradient (Fick's law) and vesicle-based active transport along polarized actin cables (Figure 1B,C).

Once a polar zone has been established, the ensuing concentration gradient on the membrane leads to a diffusive flux of proteins away from the polar zone. To maintain the polar zone, this flux on the membrane must be counteracted continually by (re-)cycling the proteins back to the polar zone via a flux from the cytosol to the membrane [15, 17, 18] or via vesicle-based transport [14, 16]. In WT cells, Cdc42-GTP recruits Bem1 from the cytosol which in turn recruits the GEF (Guanine nucleotide Exchange Factor) Cdc24 (see fig. 2.1 d.) [13, 19]. The membrane-bound Bem1-Cdc24 complex then recruits more Cdc42-GDP from the cytosol and activates it (nucleotide exchange) [16]. The hallmark and crucial element of this *mutual recruitment mechanism* is the co-localization of Cdc42 and its GEF Cdc24 [4, 15, 20–22].

Deletion of Bem1 disrupts localized Cdc42 recruitment and activation [4, 22] and thereby severely impedes the cells' ability to polarize and bud [13, 23]. *Bem1*Δ cells can be rescued by Bem1 fragments that cannot mediate mutual recruitment of Cdc42 and its GEF Cdc24, but only confer increased global (homogeneous) GEF activity by relieving Cdc24's auto-inhibition [24–27]. Even more intriguingly, in experimental evolution, *bem1*Δ mutants are reproducibly rescued by the subsequent loss of Bem3 [3], one of four known Cdc42-GAPs that catalyze the GTP-hydrolysis, i.e., switch Cdc42 into its inactive, GDP-bound state. These experimental findings suggests that there is a hidden Cdc42 polarization mechanism that is independent of GEF co-localization and is activated by either increased GEF activity or the loss of a Cdc42-GAP.

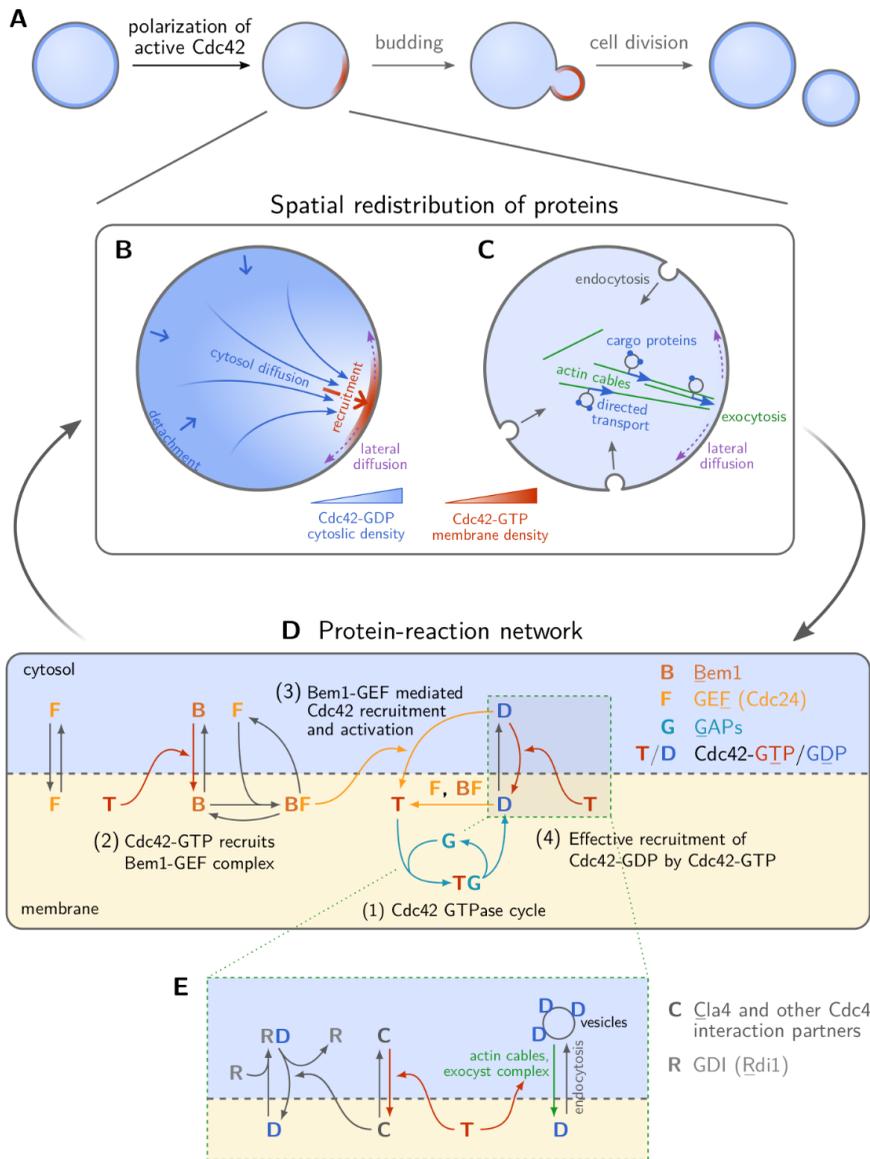


Figure 2.1. a. Starting from an initially homogenous distribution of Cdc42, a polar zone forms, marked by a high concentration of active Cdc42 on the plasma membrane. There are two pathways of directed transport in the cells: **b.** Cytosolic diffusive flux driven by a concentration gradient that is sustained by spatially separated attachment (red arrow) and detachment (blue arrow) zones; **c.** Vesicle transport (endocytic recycling) is directed along polar-oriented actin cables. Active Cdc42 directs both cytosolic diffusion (by recruiting downstream effectors that in turn recruit Cdc42) as well as vesicle transport (by recruiting Bni1 which initiates actin polymerization). **d.** Molecular interaction network around the GTPase Cdc42, involving activity regulators (GEF, GAPs), and the scaffold protein Bem1 (some components are displayed multiple times for visual clarity, not to imply a chronological order). An effective recruitment term accounts for Cdc42-recruitment to the membrane directed by Cdc42-GTP facilitated by Cdc42's interaction partners, for instance Cla4 [4–6] and Rsr1 [7]. **e.** Details of the model and the mathematical implementation are described in the Methods and Supplementary Note 1, from original paper [8]. For simplicity, we do not explicitly account for Cdc42-effector complexes. A model extension accounting for those complexes did not significantly change the results.

Here, we develop a mathematical model for the cell polarization module of budding yeast – synthesizing the insights from a large body of experimental and theoretical literature. Our theoretical analysis of this model shows that the cell-polarity module comprises multiple redundant mechanisms based on reaction–diffusion and potentially vesicle-based transport. It reveals that in addition to the Bem1-mediated mutual recruitment mechanism, a distinct and latent mechanism exists in the Cdc42-polarization machinery. Crucially, this latent mechanism requires explicit modeling of the intermediate Cdc42-GAP complex, which was not accounted for by previous models. We show that the latent mechanism operates under different constraints on the protein copy numbers than the wild-type mechanism and is activated by the loss of Bem3 which lowers the total protein copy number of GAPs. This explains how cell polarization is rescued in *bem1Δ bem3Δ* cells [3], and also reconciles the puzzling experimental findings outlined above. Moreover, we experimentally confirm the predictions of our theory on how cell polarization in various mutants can be rescued by changing the Cdc42 protein copy number. On the basis of the mechanistic understanding of the cell polarization module in budding yeast, we then propose a possible evolutionary scenario for the emergence of this self-organized cellular function. We formulate a concrete hypothesis how evolution might leverage scaffold proteins to introduce new connections in an existing network, and thus increase redundancy of mechanisms within a functional cellular module. This redundancy loosens the constraints on the module and thereby enables further evolution of its components, for instance by duplication and sub-functionalization [28].

2.2. Results

As basis for our theoretical analysis, we first need to formulate a mathematical model of the cells' Cdc42-polarization machinery that is able to explain Bem1-independent polarization. The interplay of spatial transport processes (fig. 2.1 b-c) and protein-protein interactions (fig. 2.1 d) is described in the framework of reaction–diffusion dynamics. The biochemical interaction network we propose is based on the quantitative model introduced in [17] and makes several minimal, but essential extensions to it. The model accounts for the Cdc42 GTPase cycle and the interactions between Cdc42, Bem1 and Cdc24 [15]. Extending previous models, we explicitly incorporate the transient formation of a GAP-Cdc42 complex as an intermediate step in the enzymatic interaction between GAPs and Cdc42 [29]. Explicitly accounting for the GAPs' enzyme kinetics, which was neglected in previous models [30–32] is important to account for (partial) GAP saturation in regions of high Cdc42 concentration, which will play an essential role in our findings. In addition, we include effective self-recruitment of Cdc42-GDP to the membrane which is facilitated by membrane-bound Cdc42-GTP. This effective recruitment accounts for vesicle-based Cdc42 transport along actin cables [16, 33, 34] and putative recruitment pathways mediated by Cdc42-GTP downstream effectors such as Cla4 and Gic1/2 [5, 35, 36]. A detailed description of the model, illustrated in fig. 2.1 d., and an in-depth biological motivation for the underlying assumptions are given in the supplementary notes 1 from [8].

2.2.1. The Cdc42 interaction network facilitates a latent polarization-mechanism

We first ask whether the proposed reaction–diffusion model of the Cdc42 polarization machinery can explain spontaneous polarization in the absence of Bem1, i.e. without GEF co-localization with Cdc42. To this end, we perform a linear stability analysis of the model which identifies the regimes of self-organized pattern formation. A large-scale parameter study (see supplementary notes 5 [8]) reveals that in the absence of Bem1 there is a range of protein numbers of Cdc42

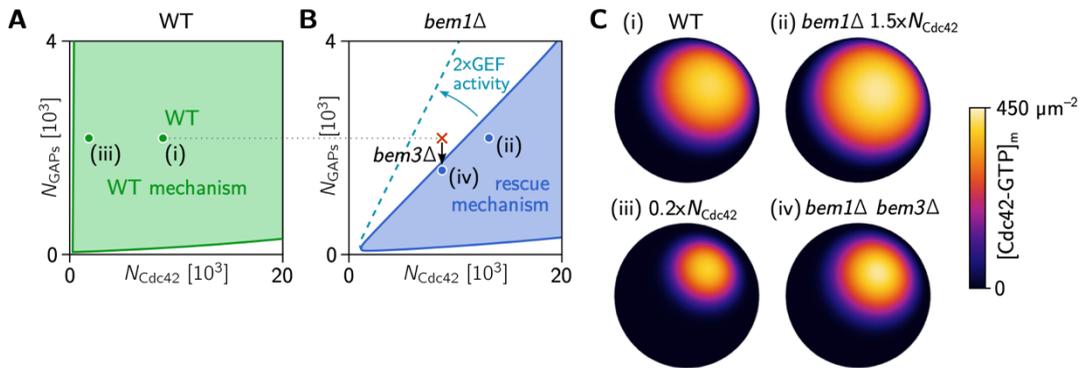


Figure 2.2. Regimes of operation of the Bem1-mediated wild-type mechanism and the latent mechanism for cell polarity. Stability diagrams as a function of GAP- and Cdc42 concentrations in presence and absence of Bem1 obtained by linear stability analysis (see supplementary notes 3 [8]) of the mathematical model for the Cdc42-polarization machinery (see fig. 2.1 and supplementary notes 2 [8]). Shaded areas indicate regimes of lateral instability, i.e. where a spontaneous polarization is possible. **A** In WT cells, the scaffold protein Bem1 is present and facilitates spontaneous polarization by a mutual recruitment mechanism that is operational in a large range of Cdc42 and GAP concentrations [15, 17]. The green point marks the Cdc42 and GAP concentrations of WT cells. **B** In the absence of Bem1, spontaneous polarization is restricted to a much smaller parameter-space region in our model, because the regime of operation of the Bem1-independent mechanism is inherently delimited by a critical ratio of GAP concentration to Cdc42 concentration. The Cdc42 and GAP concentrations of *bem1Δ* cells and *bem1Δ bem3Δ* are marked by the red cross and blue point, respectively. The experimental observation that *bem1Δ* cells do not polarize, whereas *bem1Δ bem3Δ* polarize can be used to infer a range for the critical GAP/Cdc42-concentration ratio. Increasing the GEF activity of Cdc24 increases this critical ratio (dashed blue line). **C** Snapshots from numerical simulations showing the concentration of membrane bound Cdc42-GTP in the final steady state for various mutant and copy number conditions, corresponding to Videos 1-4 (from supplementary section of [8]). (In panel (iii), the color bar represents concentrations in the range 0–200 μm^{-2}). (Model parameters were obtained by sampling for parameter sets that are consistent with the experimental findings on various mutants, as described in detail in supplementary notes 5 [8].)

and GAP where polar patterns are possible (fig. 2.2B), i.e., that there is a latent polarization mechanism. However, in contrast to the Bem1-dependent mutual recruitment mechanism (fig. 2.2A), we find that the regime of operation for this latent mechanism is more limited and requires a sufficiently low GAP/Cdc42-concentration ratio (fig. 2.2B). To validate the results from linear stability analysis, we performed numerical simulations of the full nonlinear, bulk-surface coupled reaction diffusions (see fig. 2.2C and videos 1-6 (from [8])); details described in supplementary notes 3 [8]).

What is the mechanistic cause for the constraint on the GAP/Cdc42-concentration ratio? To answer this question, we need to understand how the Cdc42-polarization mechanism works in the absence of Bem1. As emphasized above, Cdc42-polarization requires two essential features—directed transport of Cdc42 to the polar zone and localized activation of Cdc42 there. The first feature, directed transport, is accounted for in the model by effective recruitment of Cdc42-GDP to the membrane mediated by active Cdc42 (fig. 2.1d.).

2.2.2. GAP saturation can localize Cdc42 activity to the polar zone

How is the second feature, localization of Cdc42 activity to the polar zone, implemented in the absence of Bem1? Instead of directly increasing the rate of Cdc42 activation in the polar zone (via recruitment of the GEF Cdc24 by Bem1), localization of activity can also be achieved by decreasing the rate of Cdc42 *deactivation* in the polar zone and increasing it away from the polar zone. In fact, if enzyme saturation limits the net deactivation rate, a simple increase in Cdc42 density *generically* leads to a decrease of the Cdc42 deactivation rate (per Cdc42 molecule). Enzyme saturation of catalytic reactions occurs when the dissociation of the transient enzyme-substrate complex (here the GAP-Cdc42 complex) is the rate limiting step. The enzymes that are transiently sequestered in enzyme-substrate complexes are then not available to bind to further substrate molecules. Indeed, it has been shown that this is the case for GAP-catalyzed hydrolysis of Cdc42 in budding yeast [29]. Furthermore, enzyme saturation requires that a large fraction of enzymes is sequestered in enzyme-substrate complexes, i.e., that the total enzyme density is sufficiently low compared to the substrate density, as we found in the linear stability analysis (fig. 2.2B).

In summary, (partial) GAP saturation localizes Cdc42 activity to the polar zone: It decreases the deactivation rate in the polar zone, where Cdc42 density is high, relative to the remainder of the membrane, where Cdc42 density is low. This localized Cdc42 activity, in conjunction with transport of Cdc42 to the polar zone, drives spontaneous cell polarization. Interestingly, enzyme saturation of Cdc42 hydrolysis is one of the six theoretically possible mechanisms for pattern formation that were hypothesized by a generic mathematical analysis of feedback loops in GTPase cycles[37].

2.2.3. The latent polarization-mechanism explains the rescue of Bem1 deletion

The Bem1-independent rescue mechanism requires a sufficiently low GAP/Cdc42-concentration ratio to be functional (fig. 2.2B). This suggests that *bem1*Δ cells are not able to polarize because their GAP protein copy number is too high. Our model predicts that the loss of GAPs can rescue cell polarization by bringing their total protein copy number into a regime where the Bem1-independent mechanism is operational, as indicated by the arrow in fig. 2.2B. This is in accordance with evolution experiments showing that *bem1*Δ cells are reproducibly rescued by a subsequent loss-of-function mutation of the GAP Bem3 [3]. Bem3 accounts for approximately 25% of the total protein copy number of all Cdc42-GAPs [38], indicating that *bem1*Δ mutants are close to the GAP/Cdc42-ratio threshold of the Bem1-independent mechanism. This proximity of the protein copy numbers to the threshold explains why a low fraction (about 1 in 10⁵) of mutants are able to polarize and divide, after *BEM1* has been deleted [3]: Protein copy numbers vary stochastically from cell to cell such that a small fraction of cells lies in the concentration regime where the latent polarization mechanism drives spontaneous cell polarization. (For the four Cdc42 GAPs, a coefficient of variation around 0.14 for cell-to-cell copy-number variability has been reported [39]. This is on the same order of magnitude as the upper estimate of 25% for the GAP protein copy number reduction required to activate the Bem1-independent rescue mechanism, suggesting that this mechanism is operational in a fraction of *bem1*Δ cells.)

Rather than by the loss of a GAP, the GAP/Cdc42-concentration ratio could also be brought down by an increase of the Cdc42 protein copy number. Yet another option would be an increase of Cdc24's GEF activity which would increase the critical threshold in GAP/Cdc42-concentration ratio (see dashed line in fig. 2.2B). However, compared to a loss-of-function mutation, such mutations have a much smaller mutational target size and are therefore much less frequent. More-

over, one might wonder why it is specifically Bem3, rather than one of the other GAPs, that is lost to rescue the *bem1* Δ strain. Some hints to answer this outstanding question are provided by a detailed theoretical analysis of the rescue mechanism later in the section **Functional submodules of cell polarization**.

2.2.4. Copy number variation experiments confirm theoretical predictions

Based on the GAP/Cdc42-ratio constraint in the rescue mechanism, our theory makes two specific predictions: (i) Increasing the protein copy number (i.e. overexpression) of Cdc42 will rescue cell polarization of *bem1* Δ cells by invoking the Bem1-independent mechanism. (ii) Polarization of *bem1* Δ *bem3* Δ cells will break down if the protein copy number of Cdc42 is lowered compared to the WT level (fig. 2.2B).

To test these model predictions experimentally, we first constructed different yeast strains with Cdc42, labelled with sfGFP, under an inducible galactose promoter. This allows us to tune the Cdc42 protein copy number by varying the galactose concentration in the growth media [40]: a *bem1* Δ strain (yWKD069), a *bem1* Δ *bem3* Δ (yWKD070), and a modified WT strain (yWKD065) (see section 2.4). We confirmed that the sfGFP tag on our inducible Cdc42 does not significantly alter fitness (see fig. 2.6), in line with literature on viability and localization of another fluorescent Cdc42 sandwich fusion [41] in budding yeast. As a next step, we inoculated the different strains at varying galactose concentration in 96 well plates, that were placed in a plate reader to measure the cell density over time, and thereby determined the population growth rate (see section 2.4). For every galactose concentration, the growth rates are normalized to those of WT cells, with Cdc42 under its native promoter (yLL3a), grown at the same galactose concentration. In fig. 2.3A the normalized growth rates of the different mutants are plotted. As expected, WT cells grow at all galactose concentrations. In contrast, WT cells with Cdc42 under the galactose promoter (yWKD065), do not grow in the absence of Cdc42 (0% galactose concentration), since a failure to polarize severely impairs cell division and eventually leads to cell death and thus zero growth rate [13]. Our data show that the WT mechanism is rather insensitive to Cdc42 protein copy number, even for very low expression of Cdc42, in accordance with theory (Figure 2A).

Our model predicts that *bem1* Δ cells need the highest Cdc42 protein copy number to polarize, WT cells will need the least, and the *bem1* Δ *bem3* Δ cells should be in between. We indeed find that the *bem1* Δ strain (yWKD069) grows in media with 0.1% or higher galactose concentration. We inoculated these strains at lower galactose concentrations, but never observed growth for the *bem1* Δ and the *bem1* Δ *bem3* Δ strains in more than one technical replicate (out of 6 and 4 respectively) per condition (see table 2.3). We attribute the rare growth at low galactose concentrations to emergence of suppressor mutations. Therefore, we focus on comparing growth rates. There is strong and positive evidence that the *bem1* Δ *bem3* Δ grows faster than the *bem1* Δ in the 0.06, 0.1% and 0.2% galactose concentration respectively (Bayes factors 7, 131 and 6, and using interpretation qualifications from [42]). For WT cells with Cdc42 under the galactose promoter we observe a reduced growth rate at 0.01% galactose concentration but growth is only fully inhibited at 0% galactose concentration. All of the above experimental observations agree with our specific theoretical predictions.

Furthermore, we examined the influence of Cdc42 protein copy number on cell morphology (see fig. 2.3B) and viability. These experiments provide support for the conclusions from our growth assays, namely that viability increases and size (as proxy of polarization time [3, 43]) decreases with increasing protein copy number.

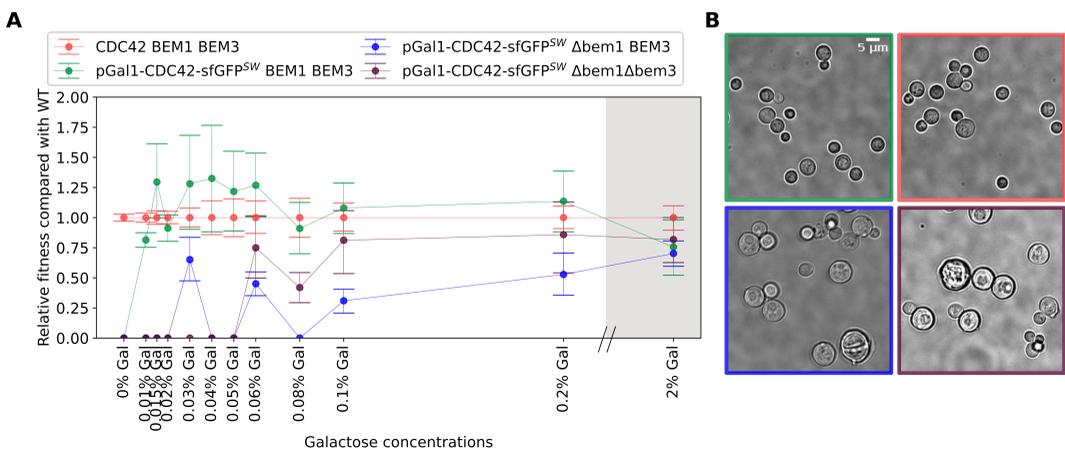


Figure 2.3. Experiments confirm theoretically predicted effect of Cdc42 protein copy number on the latent polarity-mechanism. **A** Growth rate of the different mutants (relative to the growth rate of WT cells with Cdc42 under its native promoter at that galactose concentration (red)) against the galactose concentration galactose concentration (proxy for Cdc42 protein copy number) show that higher expression of Cdc42 rescues *bem1Δ* cells and to a lesser extent *bem1Δ bem3Δ* cells; the error bar indicates the 68% credible interval, see materials and methods). **B** Micrographs of all strains in 0.06% galactose, after 24 hours of incubation. WT yeast cells with Cdc42 under the galactose and native promoter respectively are entering stationary phase and diluted 1000x (top row). The *bem1Δ* and *bem1Δ bem3Δ* cells are in log phase and diluted 100x (bottom row)

Taken together, the experimental data confirm the theoretical prediction that the Bem1-independent rescue mechanism is operational only below a threshold GAP/Cdc42-concentration ratio. In addition, we find that the Bem1-dependent WT mechanism is surprisingly insensitive to Cdc42 protein copy number, i.e., operates also at very low Cdc42 concentration. In the context of our theory, this significant difference in Cdc42 protein copy number sensitivity is explained by the qualitative difference of their principles of operation (see **The Cdc42 interaction network facilitates a latent polarization mechanism**). The WT mechanism is based on recruitment of the GEF Cdc24 to the polar zone, mediated by the scaffold protein Bem1. In contrast, the rescue mechanism crucially involves enzyme saturation of Cdc42 hydrolysis due to high Cdc42 density in the polar zone. This enzyme saturation requires a sufficiently large Cdc42 protein copy number relative to the GAP protein copy number. In the section section 2.2.6 below, we will analyze the mathematical model, and the qualitative and conceptual differences between these two mechanisms in more detail.

2.2.5. The latent rescue mechanism explains and reconciles previous experimental findings

In previous experiments, several Bem1 mutants were studied that perturb Bem1's ability to mediate co-localization of Cdc24 to Cdc42-GTP, the key feature that underlies operation of the WT mechanism [24, 25, 32, 44–46]. The observations from these experiments have remained puzzling and apparently conflicting among one another as of yet. As we show in detail in the section 2.5, the latent rescue mechanism predicted by our mathematical model explains and reconciles all of these previous experimental findings. The key insight is that the latent rescue mechanism can be activated by a global increase of GEF activity (see dashed line in fig. 2.2B). Bem1

mutants that lack the Cdc42-interaction domain but still bind to the GEF Cdc24 may provide such a global increase of GEF activity and thus rescue polarization of *bem1Δ* cells. Moreover, in accordance with optogenetics experiments [45], our mathematical model predicts that the latent Bem1-independent mechanism can also be induced outside the regime of spontaneous polarization by a sufficiently strong local perturbation of the membrane-bound GEF concentration.

2.2.6. Functional submodules of cell polarization

Cell polarization in budding yeast is a functional module based on a complex protein interaction network with Cdc42 as the central polarity protein (fig. 2.1 B-D). As we discuss next, the full network can be dissected into *functional submodules*. Here, the term functional submodule refers to a *part* of the full interaction network with a well-defined function in one or more pattern-forming mechanisms. Our theoretical analysis will reveal that an interplay of two (or more) *functional submodules* each constitutes a fully functional cell polarization mechanism. Importantly, the submodules *emerge* from the interplay of various players (components) in the biochemical interaction network and the spatial transport of proteins (by diffusion and along actin cables).

As we argued in the **Introduction**, establishment and maintenance of cell polarity requires that Cdc42-activity is localized to membrane regions with a high density of Cdc42. This can be achieved in two different ways. First, by the recruitment of the scaffold protein Bem1 to Cdc42-GTP, which in turn recruits the GEF (Cdc24) and thus localizes Cdc42 activation to the polar zone, where Cdc42 density is high (fig. 2.4A, top left). We call this the *polar activation* submodule. Second, GAP saturation in regions of high local Cdc42 densities can localize Cdc42 activity to the polar zone (fig. 2.4A, top right), as described above in the subsection **GAP saturation can localize Cdc42 to the polar zone**. The transient (partial) sequestration of GAPs in Cdc42-GAP complexes is essential for this *polar GAP saturation* submodule. The third submodule (fig. 2.4A, bottom) that we term *Cdc42 transport*, comprises various modes of Cdc42 transport towards the polar zone: vesicle transport along polarized actin cables (fig. 2.1B) and effective (self-)recruitment of Cdc42 from the cytosol. Several experiments indicate that downstream effectors of active Cdc42, such as Cla4, Gic1 and Gic2 may provide such effective recruitment in the absence of Bem1 [5, 35, 47].

These three functional submodules represent different mechanistic aspects of the Cdc42-interaction network. Each submodule is operational only under specific constraints on the biochemical properties and protein copy numbers of the involved proteins. In the following, we exploit these constraints to study the roles of the submodules in the mathematical model by disabling them one at a time. This allows us to tease apart the mechanisms that are operational under the corresponding experimental conditions. The first submodule, *polar activation*, is disabled by the knock-out of Bem1. The second submodule, *polar GAP saturation*, is suppressed if the protein copy number of GAPs is too high. Alternatively, polar GAP saturation is rendered non-operational if the dissociation rate of the GAP-Cdc42 complex is too fast, or if the free GAPs diffuse very fast making additional free GAPs readily available in the polar zone. The third submodule, *Cdc42 transport*, can be switched off by immobilizing Cdc42, i.e., suppressing its spatial redistribution. Experimentally, this has been achieved in fission yeast by fusing Cdc42 to a transmembrane protein that strongly binds to the membrane and is nearly immobile there [44].

It is worth noting that Bem1 is part of two functional submodules: Recruiting GEF to the polar zone provides *polar activation*, recruiting Cdc42 contributes to *Cdc42 transport*. While polar activation is entirely dependent on Bem1, there are several Bem1-independent modes of Cdc42 transport, including actin-based vesicle trafficking and other putative recruitment mechanisms

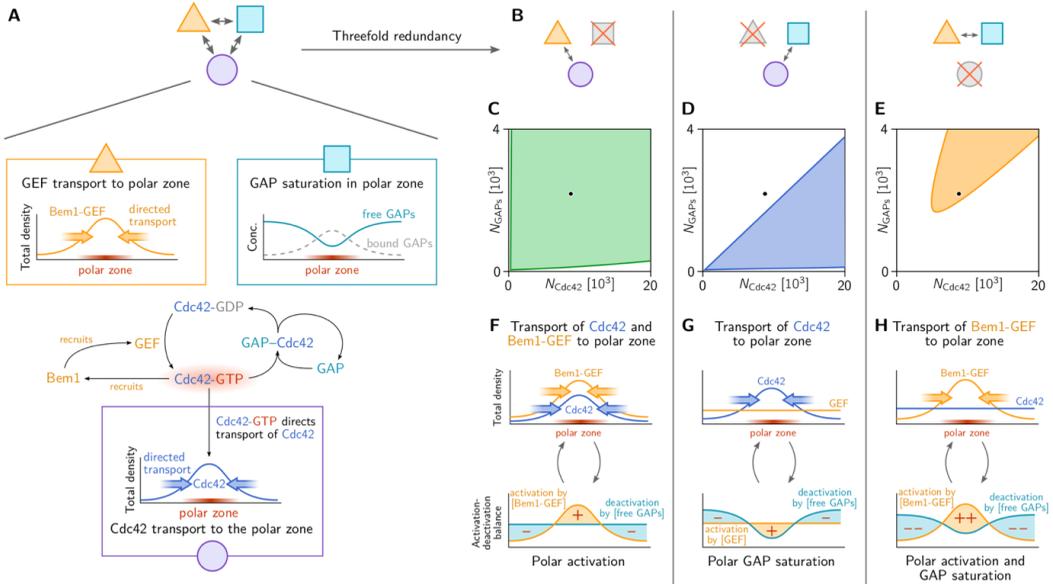


Figure 2.4. Three functional submodules constitute three distinct mechanisms of Cdc42-GTP polarization. **A** Three functional submodules of the Cdc42 interaction network contribute to the formation and maintenance of a *polar zone* (region of high Cdc42-GTP concentration, highlighted in red): Transport of Cdc42 towards the polar zone (purple circle). High Cdc42 activity can be maintained due to GAP saturation in the polar zone (teal square) and by transport of the GEF to the polar zone via the scaffold protein Bem1 (yellow triangle). **B** Combinations of pairs of these functional submodules constitute mechanisms of self-organized pattern formation. **C–E** These mechanisms are operational in different regimes of the total protein copy number of Cdc42 and GAPs. The WT mechanism (**F**) is largely insensitive to protein copy number variations (**C**) because it is based on mutual recruitment of Cdc42 and Bem1-GEF complexes, and does not depend on saturation of GAPs in the polar zone. In contrast, when the GEF is not transported to the polar zone (e.g. due to a deletion of Bem1), only GAP saturation in the polar zone maintains high Cdc42 activity there, while deactivation dominates away from the polar zone. Therefore, the polarization mechanism (**G**) is sensitive to the GAP protein copy number (**D**). **H** Remarkably, if transport of Cdc42 is suppressed, e.g. by strongly binding it to the membrane, a combination of Bem1-GEF complex recruitment and polar GAP saturation maintain a localized high Cdc42 activity even though the total density of Cdc42 is homogeneously distributed.

(fig. 2.1D.). Thus, the Cdc42 transport submodule is still operational in *bem1Δ* cells.

We next performed linear stability analysis for the full mathematical model under each of these perturbations disabling one of the submodules at a time (as described in detail in supplementary section of [8]). In each case we found that the remaining two submodules operate in concert to constitute a mechanism for spontaneous Cdc42 polarization, as illustrated in fig. 2.4B. fig. 2.4C-E shows the regime of operation of the three different mechanisms as a function of the total Cdc42 and GAP concentrations. fig. 2.4F-H illustrate the concerted interplay of directed protein-transport and regulation of Cdc42 activity (activation/deactivation) that underlie Cdc42-polarization in these three mechanisms.

Before we turn to the detailed descriptions of these mechanisms, we note that if two submodules are disabled simultaneously, the remaining submodule alone cannot facilitate pattern formation. In particular, and perhaps somewhat counterintuitively, self-recruitment of Cdc42 alone is not sufficient to drive spontaneous cell polarization [37, 48].

Wild-type mechanism: *Cdc42 transport plus polar activation*

The interplay of the Cdc42 transport submodule and the Cdc42-Bem1-Cdc24 recruitment submodule (polar activation), illustrated in fig. 2.4F, constitutes the WT mechanism that operates via mutual recruitment of Cdc42 and Bem1 [13, 16, 17]. Characteristic for this mechanism is the co-localization of Cdc24 and Cdc42-GTP in the polar zone, as observed in previous experiments [4, 45]. Other than the rescue mechanism, the mutual recruitment mechanism does not require polar GAP saturation. Therefore, it is insensitive against high concentration of GAPs, i.e., it is operational for much higher GAP/Cdc42-concentration ratios than the rescue mechanism. Furthermore, it is robust against high diffusivity of free GAPs and high catalytic rates of the GAPs (fast decay of GAP-Cdc42 complexes into free GAP and Cdc42-GDP). This implies that in mathematical models of the WT mechanism the GAPs can be accounted for *implicitly* by a constant and homogeneous hydrolysis rate, as in previous models [4, 15, 17, 49]. Notably Bem1 mediates both polar activation and Cdc42 transport (via recruitment from the cytosol) in these models.

Rescue mechanism: *Cdc42 transport plus polar GAP saturation*

The latent, Bem1-independent rescue mechanism operates by the interplay of GAP saturation in the polar zone (illustrated in fig. 2.4G) and Cdc42 transport (including effective self-recruitment via actin and/or other downstream effectors like Cla4). Characteristic for this mechanism is that it does not require co-localization of Cdc24 to Cdc42-GTP in the polar zone (see fig. 2.4G). In future experiments, this lack (or dramatic reduction) of Cdc24 polarization could serve as a clear indicator of the rescue mechanism. As explained above, the rescue mechanism relies on GAP saturation in the polar zone to maintain high Cdc42 activity there. When Cdc42 activity is maintained by lower GAP activity, we expect longer residence times of Cdc42 in the polar zone compared to WT cells. This prediction could be tested in future experiments.

GAP saturation is suppressed by either high abundance, high catalytic activity, or fast transport (by cytosolic diffusion or vesicle recycling) of the GAPs. The last constraint provides a plausible explanation why it is specifically Bem3 that needs to be deleted to rescue *bem1Δ* cells. In contrast to Rga1 and Rga2, Bem3 has been found to be highly mobile, probably because it cycles through the cytosol [50]. GAP saturation, i.e. the depletion of free GAPs in the polar zone, entails a gradient of the free GAP density towards the polar zone. A mobile GAP species like Bem3 will quickly diffuse along this gradient to replenish the free GAPs in the polar zone. This influx

relieves the GAP saturation there and thus counteracts the activation of Cdc42 in the incipient polar zone. Therefore, the loss of Bem3, rather than one of the other, less mobile GAPs, promotes the formation of a stable polar zone.

2

Polarization with immobile Cdc42: Bem1-mediated recruitment plus polar GAP saturation

The interplay of Cdc42-Bem1-Cdc24 recruitment (polar activation) and the polar GAP saturation, illustrated in fig. 2.3H, facilitates polarization of Cdc42 activity without the spatial redistribution of Cdc42's total density (blue line in fig. 2.3H, top). Instead, the proteins that are being redistributed are Bem1 and GEF. The polar zone is characterized by a high concentration of membrane-bound Bem1-GEF complexes which locally increase Cdc42 activity. Cdc42-GTP, in turn, recruits further Bem1 and GEF molecules to the polar zone. Characteristic for this mechanism is that Cdc42-GTP is polarized while the total Cdc42 density remains uniform on the membrane. Experimentally, this has been observed in fission yeast using Cdc42 fused to a transmembrane domain (Cdc42-psy1TM) that renders Cdc42 nearly immobile. The polarization machinery of fission yeast is closely related to the one of budding yeast; it operates based on the same mutual recruitment pathway with Scd1 and Scd2 taking the roles of Cdc24 and Bem1 [18]. In future experiments, it would be interesting to test whether the Cdc42-psy1TM also facilitates polarization in budding yeast (potentially in a strain with modified GAP or Cdc42 protein copy number as the regime of operation might not coincide with the WT protein copy numbers).

2.3. Discussion

2.3.1. Mechanistic understanding of the cell polarization module in budding yeast

We have discovered that multiple, redundant self-organization mechanisms coexist within the protein network underlying cell polarization in budding yeast. This explains the remarkable resilience of this module: It remains operational under many experimental (genetic) perturbations [4, 24, 25, 44, 45, 51]. While we find that the Cdc42-polarization machinery is robust against many genetic perturbations, we have put particular focus on one of its key components, Bem1, since a previous experiment has found quick and reproducible recovery from its deletion [3]. By dissecting the full cellular polarization module into *functional submodules*, we have identified three distinct mechanisms of self-organized pattern formation. Besides the wild-type mechanism relying on the colocalization of Cdc42 with its GEF via Bem1, this includes a latent and Bem1-independent rescue mechanism and a mechanism that is independent of Cdc42 redistribution. Our theory, which is compatible with published experiments, reveals that these mechanisms share many components and interaction pathways of this network. This implies that the redundancy of cell polarization is not at the level of individual components or interactions but arises on the level of the emergent function itself. If one submodule is rendered non-functional, the combination of the remaining submodules still constitutes an operational mechanism of cell polarization — if parameters, in particular protein copy numbers, are tuned to a parameter regime where these remaining submodules are operational. Redundancy hence provides adaptability — the ability to maintain function despite (genetic) perturbations. Importantly, the submodules are emergent: they involve the interplay of several network components, their biochemical interactions, and their spatial transport.

Our analysis in terms of functional submodules provides a mechanistic understanding of the polarization machinery where molecular details have been “coarse grained”. In the context of

genotype–phenotype maps, this coarse-grained description could be integrated with into a cell cycle model to address questions about epistasis [52], and eventually predict evolutionary trajectories in a population dynamics model.

Interestingly, the formation of Min-protein patterns in *E. coli* relies on the same type of mechanism as the rescue mechanism for Cdc42-polarization: self-recruitment of an ATPase (MinD) and enzyme saturation of the AAP (MinE) that catalyzes MinD’s hydrolysis and subsequent membrane dissociation [53–55]. The transient MinDE complexes play the analogous role to the Cdc42-GAP complexes here: In regions of high MinD density, MinE is sequestered in MinDE complexes, which limits the rate of hydrolysis until the complexes dissociate or additional MinE comes in by diffusion. Because MinE cycles through the cytosol, it rapidly diffuses into the polar zone where the density of free MinE is low. This diffusive influx relieves the enzyme saturation in the polar zone and eventually leads to a reversal of the MinD polarity direction. The repeated switching of MinD polarity due to redistribution of MinE is what gives rise to the Min oscillations in *E. coli*. Recently also stationary Min patterns have been observed *in vitro* [56]. Conversely, oscillatory Cdc42 dynamics are found in the fission yeast *S. Pombe* [35], and have also been indirectly observed in budding yeast mutants [49, 57].

2.3.2. The physics of self-organization imposes constraints on evolution

The fundamental question of evolutionary cell biology is “How do cells work and how did they come to be the way they are?” [58]. Our in-depth analysis of the yeast polarization machinery gives an answer to the first half of this question for a specific biological system. It also allows us to approach the second half and develop a concrete hypothesis how the Cdc42 cell-polarization machinery of budding yeast might have evolved from a more rudimental ancestral form.

Our theoretical and experimental results highlight the importance of protein copy numbers as control parameters that determine whether a mechanism of spontaneous cell polarization is operational. Phrased from a genetic perspective, the genes that code for components of the cell polarization machinery are *dosage sensitive* [59]. On the one hand, this entails that mutations of cis-regulatory elements (like promoters and enhancers) [60] can tune the protein copy numbers of proteins to the regime of operation of a specific cell-polarization mechanism and optimize the function within that regime. On the other hand, protein copy number sensitivity constrains evolution of the polarization-machinery’s components via duplication and sub-functionalization [59, 61].

One of our key findings is that the constraints on a single particular mechanism can be circumvented by the coexistence of several redundant mechanisms of self-organization that operate within the same protein-interaction network. The regimes of operation — and, hence the dosage sensitivity of specific genes — can differ vastly between these distinct mechanisms. Therefore, redundancy on the level of mechanisms allows the module’s components to overcome constraints like protein copy number sensitivity and thus promotes “evolvability” — the potential of components to acquire new (sub-)functions while maintaining the module’s original function. Previous work has shown how additional negative feedback loops can also increase the regime of operation of WT mechanism [21].

A particular example in budding yeast’s cell-polarization module where duplication and sub-functionalization might have taken place is the diversification of the different GAPs of Cdc42 in budding yeast. Bem3, Rga1, and Rga2 play individual roles in specific cellular functions, like the pheromone response pathway [50, 62], axial budding [63], and the timing of polarization [64]; see

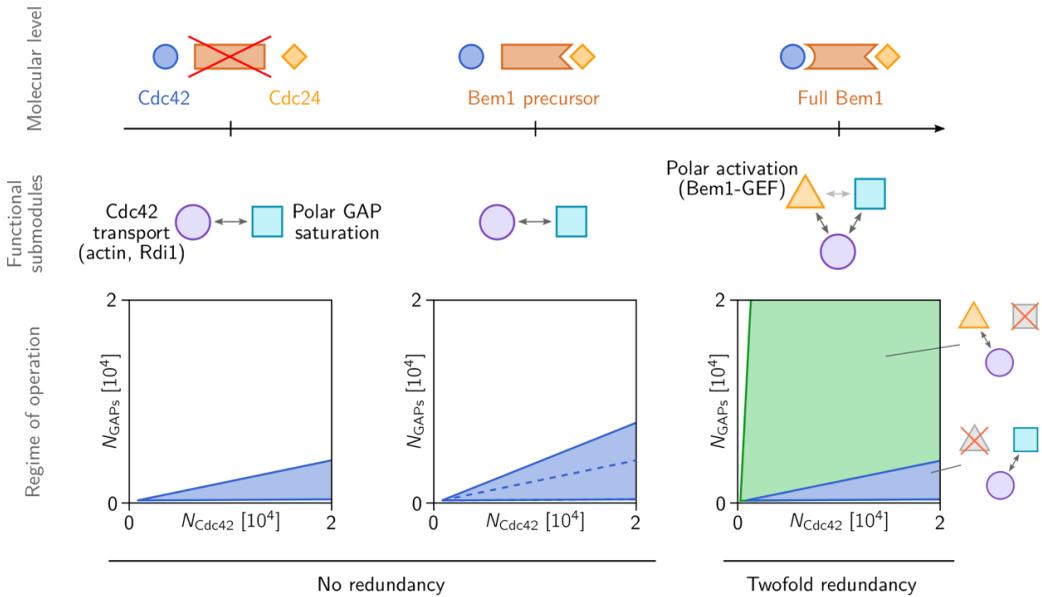


Figure 2.5. Hypothetical evolution of Bem1. (Left) The Bem1-independent “rescue” mechanism based on GAP saturation and Cdc42 transport towards membrane bound Cdc42-GTP is operational only in a limited range of the GAP/Cdc42-concentration ratios (cf. Figure 4D). (Center) a Bem1 precursor (Bem1-fragment) that binds to Cdc24 and relieves its auto-inhibition increases the range of viable GAP/Cdc42-concentration ratios and thus increases the robustness against protein copy number variations (fig. 2.2). It does, however, not change the underlying mechanism qualitatively. (Right) Domain fusion of a Cdc42-GTP-binding domain with the Cdc24-binding Bem1-precursor, leads to a new connection in the Cdc42-interaction network that leads to recruitment of Cdc24 to the polar zone. On the level of submodules, this new connection constitutes a new functional submodule that we called “polar activation” (yellow triangle). In conjunction with transport of Cdc42 towards the polar zone, polar activation gives rise to the highly robust mutual-recruitment mechanism that is operational in WT yeast (regime of operation shaded in green in the (N_D, N_G) -parameter plane; fig. 2.4C). Note that the scale on the vertical axis is chosen larger to emphasize the significantly larger regime of operation of the Bem1-mediated mechanism.

[65] for a visualization. At the origin of this diversity of GAPs is its promotion by cell-polarization mechanisms that are insensitive to GAP protein copy number, such as the Bem1-mediated WT mechanism. As we will argue below, this notion provides a concrete hypothesis about the role of scaffold proteins, like Bem1, for the evolution of functional modules that operate by the interplay of many interacting components.

2.3.3. How evolution might leverage scaffold proteins

In the context of cellular signaling processes it was suggested previously that evolution might leverage scaffold proteins to evolve new functions for ancestral proteins by regulating selectivity in pathways, shaping output behaviors and achieving new responses from preexisting signaling components [61]. Our study of the Cdc42 polarization machinery gives a perspective on how scaffold proteins may also play an important role in the evolution of intracellular self-organization. The scaffold protein Bem1 — by connecting Cdc42-GTP to Cdc42’s GEF — generates a functional submodule that contributes to self-organized Cdc42 polarization. Based on this, we propose a

hypothetical evolutionary history for Bem1, illustrated in fig. 2.5: The latent rescue mechanism is generic and rudimentary and therefore might be an ancestral mechanism of Cdc42 polarization in fungi. On this basis, Bem1 could then have evolved in a step-wise fashion: A hypothetical Bem1 precursor binding to Cdc24 but not to Cdc42-GTP might have facilitated a globally enhanced catalytic activity of Cdc24 by relieving its auto-inhibition [26, 27]. Our theory shows that such an increase of GEF activity enlarges the range of GAP/Cdc42-concentration ratios for which the latent rescue mechanism is operational. This would have entailed an evolutionary advantage by increasing the robustness of the (hypothetical) ancestral mechanism against protein copy number variations. In a subsequent step the Bem1-precursor might then have gained the Cdc42-binding domain (SH3 domain) by domain fusion [66], thus forming the full scaffold protein that connects Cdc24 to Cdc42-GTP that mediates the WT polarization mechanism (mutual recruitment of Cdc24 and Cdc42). Along this hypothetical evolutionary trajectory, the constraints on the GAP/Cdc42 protein copy number ratio and the molecular properties of the GAPs (kinetic rates, membrane affinities) would be relaxed, thereby allowing the duplication and sub-functionalization of the GAPs [61]. Given that Bem1 is highly conserved in fungi [67], and that fission yeast polarization is based on the same mutual recruitment mechanism [68, 69], this hypothetical evolutionary pathway might lie far in the past.

There are several possible routes to test our hypotheses. One possibility is the construction of phylogenetic trees for the different proteins (domains) that could inform on the order they appeared during evolution of the polarity network [70]. Another possibility is to search for species in the current tree of life which contain intermediate steps of the evolutionary trajectory. For instance species with a more ancient version of Bem1 lacking the SH3 domain, and identify the protein self-organization principles underlying polarization in these species. This is becoming a more and more realistic option, given the very large (and still expanding) number of fungal species that has been sequenced [67] and the growing interest of cell and molecular biologists to work with non-model systems [71].

On a broader perspective, we have shown how understanding the mechanistic principles underlying self-organization can provide insight into the evolution of cellular functions, a central theme in evolutionary cell biology. Specifically, we have presented a concrete example that shows how a self-organizing system might have evolve from more a rudimentary, generic mechanism that is parameter sensitive, to a specific, robust and tightly controlled mechanism by only incremental changes [72].

2.4. Materials and Methods

For the Model motivation and assumptions part of the Materials and Methods, please revise [8]. I leave in this thesis what concerns the experiments presented in Figure 3.

2.4.1. Experiments

Media. All used media has the same base with 0.69% w/v Yeast nitrogen base (Sigma) + 0.32% Amino acid mix (4x CSM) (Formedium) + 2% Raffinose (Sigma). We used different galactose concentrations, denoted as x-Gal, where x denotes the w/v % galactose percentage in the media.

Name	Genotype	Source
yLL3a	<i>MATα can1-100, leu2-3, 112, his3-11,15, ura3Δ, BUD4-S288C</i>	Ref. ³
yLL112	<i>MATα/α CAN1/ can1::P_{mfa}-HIS3, leu2-3,112/leu2-3, 112, his3-11,15/his3-11,15, ura3Δ/ura3Δ, BUD4-S288C/BUD4-S288C, BEM1/ bem1::KanMX6, BEM3/ bem3::NATMX4</i>	Ref. ³
yWKD054b	<i>MATα/α CAN1/ can1::P_{mfa}-HIS3, leu2-3,112/leu2-3, 112, his3-11,15/his3-11,15, ura3Δ/ura3Δ, BUD4-S288C/BUD4-S288C, BEM1/ bem1::KanMX6, BEM3/ bem3::NATMX4, CDC42/ CDC42::URA3-P_{gal}⁻CDC42</i>	This work
yWKD055c	<i>MATα/α CAN1/ can1::P_{mfa}-HIS3, leu2-3,112/leu2-3, 112, his3-11,15/his3-11,15, ura3Δ/ura3Δ, BUD4-S288C/BUD4-S288C, BEM1/ bem1::KanMX6, BEM3/ bem3::NATMX4, CDC42/ CDC42::URA3-P_{gal}⁻sfGFP-Cdc42^{SW}</i>	This work
yWKD065a	<i>MATα , CDC42::URA3-P_{gal}⁻sfGFP-Cdc42^{SW}, can1::P_{mfa}-HIS3, leu2-3, 112, his3-11,15, ura3Δ, BUD4-S288C</i>	This work
yWKD069a	<i>MATα, bem1:: KanMX6, CDC42::URA3-P_{gal}⁻Cdc42-sfGFP^{SW}, can1::P_{mfa}-HIS3, leu2-3, 112, his3-11,15, ura3Δ, BUD4-S288C</i>	This work
yWKD070a	<i>MATα, bem1:: KanMX6, bem3::NATMX4, CDC42::URA3- P_{gal}⁻Cdc42-sfGFP^{SW} , can1:: P_{mfa}-HIS3, leu2-3, 112, his3-11,15, ura3Δ, BUD4-S288C</i>	This work
yWKD071a	<i>MATα, CDC42::URA3-P_{gal}⁻CDC42, can1:: P_{mfa}-HIS3, leu2-3, 112, his3-11,15, ura3Δ, BUD4-S288C</i>	This work

Table 2. Strains used in this work.

Haploid strains yWKD065, yWKD069, yWKD070, yWKD071 and yWKD073 all originated from sporulation of diploids yWKD054 and yWKD055, using the lifted histidine auxotrophy for a-type haploids. Diploids yWKD054 and yWKD055 were generated by integration of plasmids pWKD010 and pWKD011 into yLL112 [3] respectively. Plasmids pWKD010 and pWKD011 consist of a pRL368 backbone [73], with a URA3 selectable marker. After amplifying this backbone without GFP, homology regions upstream and downstream of endogenous CDC42 were added with Gibson assembly, separated by an EcoRI cut site. After cutting these plasmids with EcoRI (New England

Biolabs), the homology flanks ensured the genomic integration during transformation replacing Cdc42 at its endogenous locus. Additionally, a superfolder GFP (sfGFP,[74], amino acid sequence GenBank: QLY89013.1) was added in pWKD011 with Gibson assembly between positions L134 and R135 of CDC42. This is based on previous work in *S. cerevisiae*, where a mCherry was integrated within Cdc42 [4]. We eliminated the fitness effects from *mcherry-Cdc42^{SW}* by using a superfolder GFP protein, as suggested by work in *S. pombe* (Bendezú et al., 2015). Plasmids and genomic integrations were verified by sequencing.

The assays presented in fig. 2.3 did not necessitate sfGFP, as planned localization experiments using fluorescence microscopy suffered from incomplete sfGFP degradation as documented previously in literature [75]. We tested whether the results presented, such as the growth rate differences across galactose levels, are not an artefact of adding this fluorophore, or auxotrophy differences across strains. We confirmed that the presence of the sfGFP insertion did not affect the growth rate of cells with CDC42 under the Gal promoter significantly for various galactose conditions (see Supplementary information). Moreover, medium was supplemented with four times the normal amino acid concentrations to address differences in auxotrophies between yLL3a and the other strains, and no difference was observed in maximum growth rates of YWKO065a and yLL3a in fig. 2.3.

Growth rate assays. We used a plate reader (Infinite M-200 pro, Tecan) for growth rate assays, with 96 well plates from Thermo Scientific, Nunc edge 2 96F CL, Nontreated SI lid, CAT.NO.: 267427. Rows A and H and the columns 1 and 12 were not used for measurements. We inoculated a 96-well plate with 100 μ l of medium and 5 μ l of cells (from glycerol stocks) in each well, and grew the cells in 96-well plate for 48 hours at 30 °C in a warm room. Afterwards the cells were diluted 200x into a new 96 well plate, which were then placed in the plate reader and the OD600 was measured for 48 hours using a combination of linear and orbital shaking at 36 °C. We used a home-written data analysis program in Matlab [76] to determine the log-phase doubling time for every well. The doubling time was approximated by fitting the slope of the linear regime of the log plot of the raw data. We performed at least two different experiments per condition, and per experiments we performed at least 4 technical replicates per strain/condition. The error in the growth rate plot is the 68% credible interval of the posterior distribution of these rates. The posteriors of non-WT backgrounds followed from normalization to WT rates by Monte Carlo simulations of the quotient of the original, non-normalized growth rate posteriors in a genetic background and the WT posterior in that medium. The non-normalized posteriors were calculated using the Metropolis-Hastings algorithm [77], from a rectangular prior and Student-t likelihood functions of doubling time fit estimates of all replicates in that medium. The standard errors of individual estimates come from the standard error of the slope parameter resulting from weighted least squares (WLS) on a moving window per OD curve, using an instrument error proxy for the WLS weights. The standard errors of individual estimates are corrected for overdispersion by the average modified Birge ratio [78] across media for WT.

Cell density visualization during growth rate assay. The microscopy images were taken with a Nikon Eclipse Ti-E inverted microscope with an oil immersion 60x objective, and 1.40 of numerical aperture. The images were artificially 1.5x more magnified. We used a 96 black multiwell plates compliant to the SBS (Society for Biomolecular Screening) standard-format with cover glass bottoms made from borosilicate glass. Cells were incubated using the first part of the growth rate assay protocol (in the plate reader at 30°C for 48h). Then, they were diluted 100X to a new plate and incubate at 36°C for 24 hours, before they reached complete saturation. The cells

were diluted 1000x for the WT backgrounds for all galactose concentrations, and 100x for non-WT backgrounds for galactose concentrations greater than 0.05%. The media used for incubating and diluting the cells was 4xCSM+2% Raffinose with the respective galactose concentrations, for each strain.

Cell size quantification. All microscopy images were taken with an Olympus IX81 inverted microscope equipped with Andor revolution and Yokogawa CSU X1 modules. We used a 100x oil objective. The acquisition software installed is Andor iQ3. The CG imaging plates were from Zell-Kontakt. They are black multiwell plates compliant to the SBS (Society for Biomolecular Screening) standard-format with cover glass bottoms made from borosilicate glass. Cells were grown in an overnight culture in CSM +2% Raffinose +2% Galactose media, without reaching saturation. On the next day, three washing steps with CSM+2% Raffinose were performed and subsequently the cells were re-suspended in the desired media of 0%, 0.06% and 0.1% Galactose. To obtain cell populations at all galactose concentrations, we first incubated all strains in 2% galactose concentration, where Cdc42 is highly overexpressed, such that also *bem1*Δ cells are able to efficiently polarize. After 15 hours of incubation in 2% galactose concentration, we exchanged the medium to the desired galactose concentration. After 24 hours, we observed the cells with light microscopy. After 24 hours leftover Cdc42 from the initial 2% galactose concentration incubation is (very low due to degradation and dilution (Cdc42 half-life is about 8 hours [79]) . From these images, we determined the average cell radius of the cells in the population. Note that all of them contain the same base media: CSM+2% Raffinose. Afterwards the cells were incubated for 8 hours at 30°, followed by an imaging session, and subsequently incubated for another 16 hours after which another imaging sessions was performed. We performed three independent experiments for each galactose concentration.

Microscopy data analysis. We performed bright field microscopy assays to monitor the cell size across different levels of Cdc42 in different genetic backgrounds. With ImageJ we manually determined the perimeter of the individual cells by fitting the live cells to a circle with the Measure tool. We performed three independent experiments per condition and per strain. In addition, we visually checked how many of the cells were alive and how many were dead based on their morphology. We observe what is called accidental cell death [80] upon inducing a very low Cdc42 protein copy number regulated by the Gal promoter. This type of cell deaths shows very distinctive phenotype associated to necrosis, namely: disintegration of cell structure and plasma membrane rupture. Once we observe this phenotype in our cells, we classify them as dead. The error bar on the fraction of dead cells as well as of the average cell radius, is calculated as the standard error over the total number of analyzed cells.

2.5. Appendix

2.5.1. Growth frequency assay (Fig. 3)

In the growth assay shown in Fig. 3 growth rates were determined for various genotypes/galactose concentration combinations. However, not all combinations exhibit the same frequency of growth. Table 2.2 shows how often growth occurred within the total number of replicates performed.

Error bars in Fig. 3 show the 68% confidence interval for the mean. Large error bars result from two effects: (i) For *Gal1-sfGFP-Cdc42 BEM1 BEM3*, they originate from the technical replicate variation, i.e., noise across runs/wells. (ii) For the *Δbem1* at low galactose concentrations, there is very infrequent growth. For instance, at 0.03% there was growth only in one well. In this case,

% Gal	WT*			GFP*			M1*			M2*		
	TG	#R	#E	TG	#R	#E	TG	#R	#E	TG	#R	#E
0	10	10	4	0	4	2	0	6	2	0	4	2
0.01	10	10	4	4	4	2	0	6	2	0	4	2
0.015	10	10	4	4	4	2	0	6	2	0	4	2
0.02	10	10	4	4	4	2	0	6	2	0	4	2
0.03	10	10	4	4	4	2	1	6	2	0	4	2
0.04	10	10	4	4	4	2	0	6	2	0	4	2
0.05	10	10	4	4	4	2	0	6	2	0	4	2
0.06	10	10	4	4	4	2	1	6	2	1	4	2
0.08	10	10	4	4	4	2	0	6	2	1	4	2
0.10	10	10	4	4	4	2	4	6	2	1	4	2
0.20	9	10	4	4	4	2	6	6	2	4	4	2
2.00	8	8	3	2	2	1	6	6	2	2	2	1

Table 2.2. Number of wells where growth is observed for various genotypes and conditions of the data underlying Fig. 3. TG indicates the number of runs (#R) that grew across #E experiments.* Strain abbreviations: WT: *CDC42*, GFP: *p_{GAL1}-CDC42-sfGFP^{SW}*, M1: *bem1Δ p_{GAL1}-CDC42-sfGFP^{SW}*, M2: *bem1Δ bem3Δ p_{GAL1}-CDC42-sfGFP^{SW}*.

the variance of only one Student t likelihood, which depends on the fitting error corrected for overdispersion, is the main determinant for the width of the credible interval, so this makes this interval a bit wide. However, the poor fitness of this background will likely still yield large credible intervals even with a large number of replicates, as its low fitness also makes it vulnerable for suppressor sweeps of the population. The stochasticity in growth at low galactose concentrations for *Δbem1* backgrounds is also seen in the other data points (see Table 2.3).

2.5.2. Determination of fitness effects of sfGFP-CDC42 sandwich fusion

Additional growth rate assays were performed to determine possible growth rate defects of sfGFP tagging of Cdc42. These were taken in a Biotek Epoch™ 2 Microstrain Spectrophotometer strain reader, a different type compared to that of the assay of Fig. 3. All strain backgrounds contained *BEM1* and *BEM3*. This assay compares the relative performance of a *p_{GAL1}-Cdc42-sfGFP^{SW}* strain compared to the *p_{GAL1}-CDC42* strain within this experiment. From Fig. 2.6, we can see that growth with and without sfGFP occurs at similar rates. The growth rate deteriorates notably when reducing induction for the strains with galactose-dependent Cdc42 production. However, we also witnessed that when fitness is low, there is a realistic risk that late in the measurements a relatively fast-growing population emerges in our wells. We suspect that mutants carrying a suppressor, presumably concerning the galactose induction system, arise which sweep the population. Therefore, we apply a coarse filter in an at-tempt to remove these, at least partially. For this purpose, we fit all optical density (OD) curves with a sigmoidal curve, which allows us to infer to initial OD. As during incubation colonies are growing to saturation to an OD of around 1, we expect an initial OD of about 0.01. Consequently, those wells that suggest an inferred OD at time = 0 that is more than an order of magnitude lower (< 0.001), are discarded. We confirm that we are not too relaxed with our threshold as on occasion a WT well is discarded which is unlikely to have been swept by a mutant. Indeed, we see the galactose-dependent strains have higher

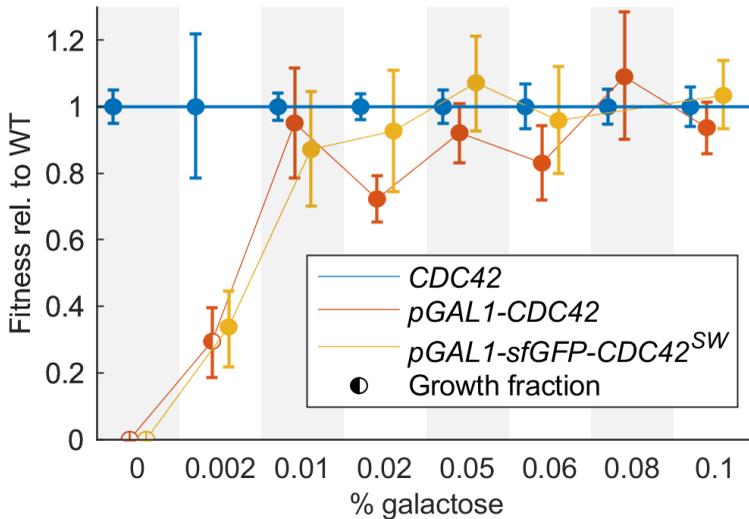


Figure 2.6. Normalized fitness data for different CDC42 variants (indicated by colors) with suppressor filtering. All backgrounds are with *BEM1* and *BEM3*. Markers indicate the mean of the posterior probability distribution for the fitness and error bars mark the 95% credible interval. Pie chart marker shows how frequently growth occurs across replicates (filled circle corresponding to 100%). The number of experiments per strain/condition pair is given in Supplementary Table 2.4.

rates of suspected suppressor growth (see Table 2.3) and we see the bulk of the suppressors are detected when fitness is low (see also Fig. 2.6). It is important to note that while this particularly cleans the growth rate data for low galactose concentrations, this filtering is usually inconsequential for the question whether sfGFP addition influences fitness (Table 2.3). Using as our metric the Bayes factor (posterior odds ratio) that the strain which is observed to grow faster is actually faster against the opposite statement, we require a value > 10 for strong evidence [81]. After filtering, there is never strong evidence for significant growth rate difference between having sfGFP or not.

2.5.3. Size and viability assays under variable expression of CDC42

To further substantiate that the differences in population growth rates are directly caused by the ability of cells to polarize, rather than for example pleiotropic changes in another cell cycle phase, we measured the cell radius using light microscopy (Fig. 2.8A). All microscopy was performed with a Nikon Eclipse Ti-E inverted microscope with an oil immersion 60X objective with NA 1.40 and refractive index of 1.51. The software used for data collection was Nis-Elements Advanced research version 4.51. Data collection of brightfield microscopy images were acquire over three different galactose concentrations (0% , 0.06% and 0.1%) after 24h of incubation of the strains: ywkd065, ywkd069, ywkd070 and yll3a.

It was previously shown that the cell radius correlates linearly with the time it takes for cells to polarize [3, 43]: cells that take longer to polarize are on average larger than cells that polarize fast because yeast cells continue to grow during polarity establishment, allowing us to use the cell radius as a proxy for the polarization time. Additionally, we verified that, at low Cdc42 dosages, cells cannot polarize at all and thus die. Consistent with the population growth data, we observed

% Gal	<i>CDC42</i>				$p_{GAL1-CDC42}$				$p_{GAL1-CDC42-sfGFP^{SW}}$			
	#D	TG	#R	#E	#D	TG	#R	#E	#D	TG	#R	#E
0	4	8	8	6	12	0	18	9	9	0	3	4
0.002	0	2	2	1	0	1	2	1	0	1	1	1
0.01	2	10	10	6	10	2	2	6	0	2	2	2
0.02	1	11	11	6	6	6	6	6	0	2	2	2
0.05	1	9	9	5	3	7	7	5	0	1	1	1
0.06	0	8	8	4	3	7	7	5	1	3	3	3
0.08	0	4	4	2	3	1	1	2	-	-	-	-
0.10	2	10	10	6	3	19	19	9	0	8	8	4

Table 2.3. Number of wells where growth is observed for the *CDC42* variants in the growth assay data with suppressor filtering. All backgrounds are with *BEM1* and *BEM3*. The column #D denotes how many wells with growth have been marked as potentially from suppressors and are thus discarded from the analysis, the remainder of the growing wells are in the columns named times (TG) growth and runs (#R). #E denotes the number of experiments that comprise all runs (#R) including those discarded ones.

% Gal	0.002	0.01	0.02	0.05	0.06	0.1
Bayes factor without filtering	2	22	7	4	7	4
Bayes factor with filtering	2	2	7	5	3	5

Table 2.4. Posterior odds ratio (Bayes factor) per medium condition of the probability that the strain with observed faster growth (between the $p_{GAL1-CDC42}$ and $p_{GAL1-CDC42-sfGFP^{SW}}$ strains) is actually faster, divided by the probability that the opposite is true. We distinguish the cases where we aim to filter suppressor growth away from our data sets and the case where we do not.

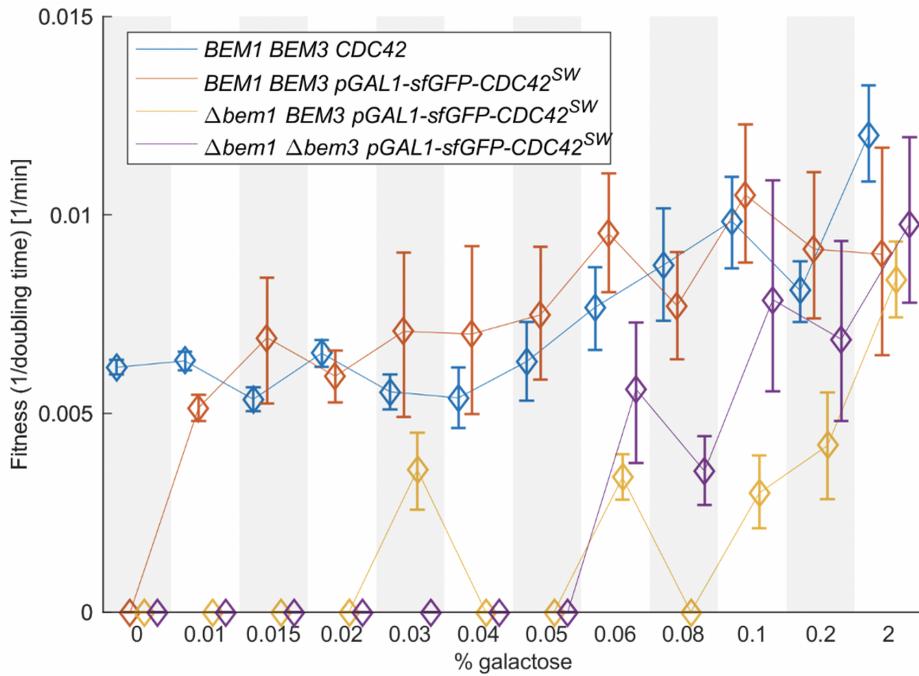


Figure 2.7. Non-normalized growth rates of the different mutant strains against galactose concentration (relates to Fig. 3 in the main text). Markers show the mean of the posterior probability distribution for the fitness and the error bars indicate the 68% credible interval; see Methods section in the main text. The number of experiments per strain/condition pair is given in Supplementary Table 2.3.

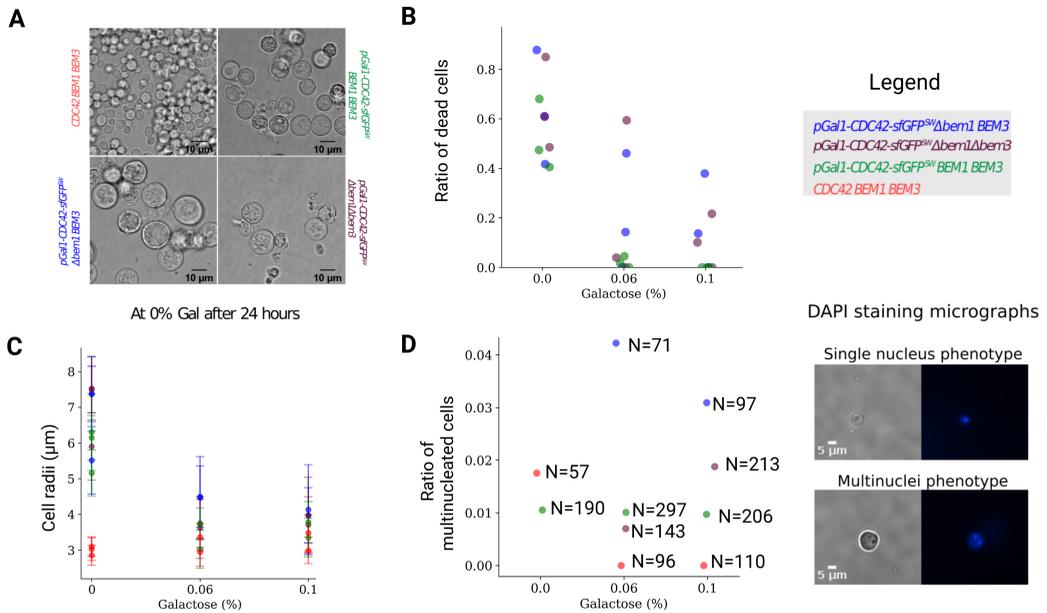


Figure 2.8. Morphology and viability for WT and mutant strains with pGAL1-CDC42-sfGFP^{sw} at various induction levels. **A** Microscopy images show the morphology of dead and alive cells after 24 h at 0% galactose concentration, resulting in a Cdc42 dosage that approximates zero; scale bar indicates 10 μ m. **B** Ratio of dead cells of the relevant mutants over different galactose concentrations. The data contains information from three independent experiments. **C** Cell radii of the relevant mutants over different galactose concentrations. Each point represents the arithmetic mean of the cell radii per technical replicate (every independent experiment). Error bars represent the standard deviation of each dataset. **D** Ratio of multinucleated cells of the relevant mutants over different galactose concentrations. The data contains information from one experiment per genotype.

that after 24 hours at 0% galactose concentration, for every genetic background where Cdc42 is under the galactose promotor, the vast majority of cells are not able to polarize or polarize very slowly, because they are either dead (Fig. 2.8A–C) or very large (Fig. 2.8A, D). We also confirm that the average cell radius (and thus the polarization time) and death rate of cells with Cdc42 under its native promotor are not affected by the galactose concentration (Fig. 2.8C, D in red). At 0.06% galactose concentration, $\Delta bem1\Delta bem3$ cells' radii (and thus polarization times) are closer to WT cell radii than those of $\Delta bem1$ cells. This agrees with the population growth data. And at 0.1% galactose concentration, the average cell radius for live cells for all mutants were approximately equal to the average WT cell radius (Fig. 2.8D). Interestingly, after 24 hours at 0% galactose concentration, WT cells with Cdc42 under the galactose promotor are still polarizing faster than the $\Delta bem1$ and the $\Delta bem1\Delta bem3$ cells, as indicated by their smaller average cell radius (Fig. 2.8D). This observation confirms our above observation that a very small number of Cdc42 molecules is sufficient for WT cells to polarize and thus for the WT mechanism to be operational.

2.5.4. Multinucleated and dead cells

Cells that do not polarize are predicted to have nuclear divisions but no cellular division and therefore, we hypothesize that cells with polarization defects are more likely to be multinucleated.

Multinucleated cells.

We measured the percentage of multinucleated cells for the different genetic backgrounds and galactose concentrations using DAPI staining. The DAPI staining protocol was taken from [82]. Two experiments failed due to two causes: too little staining for the amount of analysed cells and DAPI stain saturation for taking a long time (> 30 min) from the staining to the microscope. In the last situation most of the cells were stained and not only the nuclei. Samples were imaged in a cover slip right after the DAPI staining. The emission wavelength was 450 nm and the excitation wavelength was 395 nm. The exposure time was 70ms and the laser power was 1%. Data analysis was performed with the Cell Counter plugin from ImageJ (v1.53t). Each cell was labelled to count the total number of cells in each frame. Further, each cell with two or more nuclei and the dead cells were differently labelled. After DAPI staining we do not observe any cells at 0% galactose concentration for the $\Delta bem1$ and $\Delta bem1\Delta bem3$ cells that were alive before staining (dead cells are permeable and overstain for DAPI), and thus we also do not observe any multinucleated cells. These results also suggest that we underestimate the number of dead cells when we use bright field microscopy as in Fig. 2.8. The wild-type backgrounds with and without the galactose promoter show a low percentage of multinucleated cells. As the galactose concentration goes up the fraction of dead cells for the $\Delta bem1$ and $\Delta bem1\Delta bem3$ mutants goes down. The number of detected multinucleated cells remains low. We interpret that our results are consistent with the hypothesis that multinucleated cells die more rapidly in stationary phase than single nucleated cells and therefore are hard to detect in this assay.

Cell radii.

We used the “oval tool” to encircle every cell and measure its perimeter, using the “Analyse” → “Measure” tool. From the perimeters p , the cell radii were computed using the formula $r = p/(2\pi)$.

Dead cells.

To identify dead cells, we looked at the following morphological features [83]: cell shrinkage, swelling of organelles, nuclear fragmentation and plasma membrane blebbing.

2.5.5. Acknowledgments

We thank Felix Meigel and Marit Smeets for their pioneering experimental work. We thank Eelco Tromer for his advice on phylogenetics. We thank Daniel Needleman and Andrew Goryachev for critical reading of the manuscript. EF acknowledges support by the Excellence Cluster ORIGINS, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311. LL and LIC acknowledge support from the Netherlands Organization for Scientific Research (NWO) through a VIDI grant (016.Vidi.171.060). LL and WD acknowledge support from the Netherlands Organization for Scientific Research (NWO/OCW), as part of the Gravitation Program: Frontiers of Nanoscience. LL acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (*Grant agreement No. 758132*).

The authors declare no competing interests.

2.5.6. Data Availability Statement

Experimental data generated for this study are available in the 4TU repositories at <http://doi.org/10.4121/c.5123249.v1> and <http://doi.org/10.4121/12961541.v1>

2.5.7. Code Availability Statement

Mathematica code for linear stability analysis of the mathematical model and COMSOL Multiphysics simulation files are provided in the GitHub repository <https://github.com/f-brauns/yeast-polarity-LSA>. This repository also contains parameter sets filtered from large scale parameter space sampling. The Matlab and Python scripts used to analyze experimental data are provided in the 4TU repository at <http://doi.org/10.4121/13042796>.

Bibliography

1. Glazenburg, M. M. & Laan, L. Complexity and self-organization in the evolution of cell polarization. *Journal of cell science* **136**, jcs259639 (2023).
2. Diepeveen, E. T., de la Cruz, L. I. & Laan, L. Evolutionary dynamics in the fungal polarization network, a mechanistic perspective. *Biophysical Reviews* **9**, 375–387 (2017).
3. Laan, L., Koschwanetz, J. H. & Murray, A. W. Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *Elife* **4**, e09638 (2015).
4. Woods, B., Kuo, C.-C., Wu, C.-F., Zyla, T. R. & Lew, D. J. Polarity establishment requires localized activation of Cdc42. *Journal of Cell Biology* **211**, 19–26 (2015).
5. Tiedje, C., Sakwa, I., Just, U. & Höfken, T. The rho gdi rdi1 regulates rho gtpases by distinct mechanisms. *Molecular biology of the cell* **19**, 2885–2896 (2008).
6. Chen, G.-C., Kim, Y.-J. & Chan, C. S. The Cdc42 GTPase-associated proteins Gic1 and Gic2 are required for polarized cell growth in *Saccharomyces cerevisiae*. *Genes & development* **11**, 2958–2971 (1997).
7. Martin, S. G. Spontaneous cell polarization: Feedback control of Cdc42 GTPase breaks cellular symmetry. *Bioessays* **37**, 1193–1201 (2015).
8. Brauns, F. *et al.* Redundancy and the role of protein copy numbers in the cell polarization machinery of budding yeast. *Nature Communications* **14**, 6504 (2023).
9. Kang, P. J., Sanson, A., Lee, B. & Park, H.-O. A GDP/GTP exchange factor involved in linking a spatial landmark to cell polarity. *Science* **292**, 1376–1378 (2001).
10. Marston, A. L., Chen, T., Yang, M. C., Belhumeur, P. & Chant, J. A localized GTPase exchange factor, Bud5, determines the orientation of division axes in yeast. *Current Biology* **11**, 803–807 (2001).
11. Kozminski, K. G. *et al.* Interaction between a Ras and a Rho GTPase couples selection of a growth site to the development of cell polarity in yeast. *Molecular biology of the cell* **14**, 4958–4970 (2003).
12. Bi, E. & Park, H.-O. Cell polarization and cytokinesis in budding yeast. *Genetics* **191**, 347–387 (2012).
13. Irazoqui, J. E., Gladfelter, A. S. & Lew, D. J. Scaffold-mediated symmetry breaking by Cdc42p. *Nature cell biology* **5**, 1062–1070 (2003).
14. Wedlich-Soldner, R., Altschuler, S., Wu, L. & Li, R. Spontaneous cell polarization through actomyosin-based delivery of the Cdc42 GTPase. *Science* **299**, 1231–1235 (2003).
15. Goryachev, A. B. & Pokhilko, A. V. Dynamics of Cdc42 network embodies a Turing-type mechanism of yeast cell polarity. *FEBS letters* **582**, 1437–1443 (2008).
16. Freisinger, T. *et al.* Establishment of a robust single axis of cell polarity by coupling multiple positive feedback loops. *Nature communications* **4**, 1807 (2013).

17. Klünder, B., Freisinger, T., Wedlich-Söldner, R. & Frey, E. GDI-mediated cell polarization in yeast provides precise spatial and temporal control of Cdc42 signaling. *PLoS computational biology* **9**, e1003396 (2013).
18. Chiou, J.-g., Balasubramanian, M. K. & Lew, D. J. Cell polarity in yeast. *Annual review of cell and developmental biology* **33**, 77–101 (2017).
19. Bose, I. *et al.* Assembly of scaffold-mediated complexes containing Cdc42p, the exchange factor Cdc24p, and the effector Cla4p required for cell cycle-regulated phosphorylation of Cdc24p. *Journal of Biological Chemistry* **276**, 7176–7186 (2001).
20. Butty, A.-C. *et al.* A positive feedback loop stabilizes the guanine-nucleotide exchange factor Cdc24 at sites of polarization. *The EMBO journal* (2002).
21. Howell, A. S. *et al.* Singularity in polarization: rewiring yeast cells to make two buds. *Cell* **139**, 731–743 (2009).
22. Kozubowski, L. *et al.* Symmetry-breaking polarization driven by a Cdc42p GEF-PAK complex. *Current Biology* **18**, 1719–1726 (2008).
23. Chenevert, J., Corrado, K., Bender, A., Pringle, J. & Herskowitz, I. A yeast gene (BEM1) necessary for cell polarization whose product contains two SH3 domains. *Nature* **356**, 77–79 (1992).
24. Smith, S. E. *et al.* Independence of symmetry breaking on Bem1-mediated autocatalytic activation of Cdc42. *Journal of Cell Biology* **202**, 1091–1106 (2013).
25. Grinhagens, S. *et al.* A time-resolved interaction analysis of Bem1 reconstructs the flow of Cdc42 during polar growth. *Life Science Alliance* **3** (2020).
26. Shimada, Y., Wiget, P., Gulli, M.-P., Bi, E. & Peter, M. The nucleotide exchange factor Cdc42p may be regulated by auto-inhibition. *The EMBO journal* **23**, 1051–1062 (2004).
27. Rapali, P. *et al.* Scaffold-mediated gating of Cdc42 signalling flux. *Elife* **6**, e25257 (2017).
28. Santoshkumar Magadam, S. M., Urbi Banerjee, U. B., Priyadharshini Murugan, P. M., Doddabhimappa Gangapur, D. G. & Rajasekar Ravikesavan, R. R. Gene duplication as a major force in evolution. (2013).
29. Zhang, B., Wang, Z.-X. & Zheng, Y. Characterization of the interactions between the small GTPase Cdc42 and its GTPase-activating proteins and putative effectors: Comparison of kinetic properties of Cdc42 binding to the Cdc42-interactive domains. *Journal of Biological Chemistry* **272**, 21999–22007 (1997).
30. Khalili, B., Merlini, L., Vincenzetti, V., Martin, S. G. & Vavylonis, D. Exploration and stabilization of Ras1 mating zone: A mechanism with positive and negative feedbacks. *PLoS computational biology* **14**, e1006317 (2018).
31. Okada, S. *et al.* Daughter cell identity emerges from the interplay of Cdc42, septins, and exocytosis. *Developmental cell* **26**, 148–161 (2013).
32. Howell, A. S. *et al.* Negative feedback enhances robustness in the yeast polarity establishment circuit. *Cell* **149**, 322–333 (2012).
33. Layton, A. T. *et al.* Modeling vesicle traffic reveals unexpected consequences for Cdc42p-mediated polarity establishment. *Current Biology* **21**, 184–194 (2011).

34. Slaughter, B. D., Das, A., Schwartz, J. W., Rubinstein, B. & Li, R. Dual modes of cdc42 recycling fine-tune polarized morphogenesis. *Developmental cell* **17**, 823–835 (2009).
35. Das, A. *et al.* Flippase-mediated phospholipid asymmetry promotes fast Cdc42 recycling in dynamic maintenance of cell polarity. *Nature cell biology* **14**, 304–310 (2012).
36. Daniels, C. N., Zyla, T. R. & Lew, D. J. A role for Gic1 and Gic2 in Cdc42 polarization at elevated temperature. *PLoS one* **13**, e0200863 (2018).
37. Goryachev, A. B. & Leda, M. Many roads to symmetry breaking: molecular mechanisms and theoretical models of yeast cell polarity. *Molecular biology of the cell* **28**, 370–380 (2017).
38. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature methods* **11**, 319–324 (2014).
39. Chong, Y. T. *et al.* Yeast proteome dynamics from single cell imaging and automated analysis. *Cell* **161**, 1413–1424 (2015).
40. Yocum, R. R., Hanley, S., West Jr, R. & Ptashne, M. Use of lacZ fusions to delimit regulatory elements of the inducible divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **4**, 1985–1998 (1984).
41. Lu, M. S. & Drubin, D. G. Cdc42 GTPase regulates ESCRTs in nuclear envelope sealing and ER remodeling. *Journal of Cell Biology* **219**, e201910119 (2020).
42. Kass, R. E. & Raftery, A. E. Bayes factors. *Journal of the American Statistical Association* **90**, 773–795 (1995).
43. Allard, C. A., Decker, F., Weiner, O. D., Toettcher, J. E. & Graziano, B. R. A size-invariant bud-division timer enables robustness in yeast cell size control. *PLoS one* **13**, e0209301 (2018).
44. Bendezú, F. O. *et al.* Spontaneous Cdc42 polarization independent of GDI-mediated extraction and actin-based trafficking. *PLoS biology* **13**, e1002097 (2015).
45. Witte, K., Strickland, D. & Glotzer, M. Cell cycle entry triggers a switch between two modes of Cdc42 activation during yeast polarization. *Elife* **6**, e26722 (2017).
46. Woods, B. *et al.* Parallel actin-independent recycling pathways polarize Cdc42 in budding yeast. *Current Biology* **26**, 2114–2126 (2016).
47. Kang, P. J., Miller, K. E., Guegueniat, J., Beven, L. & Park, H.-O. The shared role of the Rsr1 GTPase and Gic1/Gic2 in Cdc42 polarization. *Molecular biology of the cell* **29**, 2359–2369 (2018).
48. Altschuler, S. J., Angenent, S. B., Wang, Y. & Wu, L. F. On the spontaneous emergence of cell polarity. *Nature* **454**, 886–889 (2008).
49. Kuo, C.-C. *et al.* Inhibitory GEF phosphorylation provides negative feedback in the yeast polarity circuit. *Current Biology* **24**, 753–759 (2014).
50. Mukherjee, D. *et al.* Bem3, a Cdc42 GTPase-activating protein, traffics to an intracellular compartment and recruits the secretory Rab GTPase Sec4 to endomembranes. *Journal of cell science* **126**, 4560–4571 (2013).
51. Brown, J. L., Jaquenoud, M., Gulli, M.-P., Chant, J. & Peter, M. Novel Cdc42-binding proteins Gic1 and Gic2 control cell polarity in yeast. *Genes & development* **11**, 2972–2982 (1997).

52. Daalman, W., Sweep, E. & Laan, L. A tractable bottom-up model of the yeast polarity genotype-phenotype map for evolutionary relevant predictions (2020).
53. Huang, K. C., Meir, Y. & Wingreen, N. S. Dynamic structures in *Escherichia coli*: spontaneous formation of MinE rings and MinD polar zones. *Proceedings of the National Academy of Sciences* **100**, 12724–12728 (2003).
54. Halatek, J. & Frey, E. Highly canalized MinD transfer and MinE sequestration explain the origin of robust MinCDE-protein dynamics. *Cell reports* **1**, 741–752 (2012).
55. Halatek, J., Brauns, F. & Frey, E. Self-organization principles of intracellular pattern formation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170107 (2018).
56. Glock, P. *et al.* Stationary patterns in a two-protein reaction-diffusion system. *ACS Synthetic Biology* **8**, 148–157 (2018).
57. Ozbudak, E. M., Becskei, A. & Van Oudenaarden, A. A system of counteracting feedback loops regulates Cdc42p activity during spontaneous cell polarization. *Developmental cell* **9**, 565–571 (2005).
58. Lynch, M. *et al.* Evolutionary cell biology: two origins, one objective. *Proceedings of the National Academy of Sciences* **111**, 16990–16994 (2014).
59. Papp, B., Pál, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194–197 (2003).
60. Wittkopp, P. J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics* **13**, 59–69 (2012).
61. Conant, G. C. & Wolfe, K. H. Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics* **9**, 938–950 (2008).
62. Stevenson, B. J. *et al.* Mutation of RGA1, which encodes a putative GTPase-activating protein for the polarity-establishment protein Cdc42p, activates the pheromone-response pathway in the yeast *Saccharomyces cerevisiae*. *Genes & development* **9**, 2949–2963 (1995).
63. Tong, Z. *et al.* Adjacent positioning of cellular structures enabled by a Cdc42 GTPase-activating protein-mediated zone of inhibition. *The Journal of cell biology* **179**, 1375–1384 (2007).
64. Knaus, M. *et al.* Phosphorylation of Bem2p and Bem3p may contribute to local activation of Cdc42p at bud emergence. *The EMBO journal* **26**, 4501–4513 (2007).
65. Daalman, W. K.-G., Sweep, E. & Laan, L. The Path towards Predicting Evolution as Illustrated in Yeast Cell Polarity. *Cells* **9**, 2534 (2020).
66. Farr, A. D., Remigi, P. & Rainey, P. B. Adaptive evolution by spontaneous domain fusion and protein relocalization. *Nature ecology & evolution* **1**, 1562–1568 (2017).
67. Diepeveen, E. T., Gehrman, T., Pourquié, V., Abeel, T. & Laan, L. Patterns of conservation and diversification in the fungal polarization network. *Genome Biology and Evolution* **10**, 1765–1782 (2018).
68. Lamas, I., Merlini, L., Vještica, A., Vincenzetti, V. & Martin, S. G. Optogenetics reveals Cdc42 local activation by scaffold-mediated positive feedback and Ras GTPase. *PLoS biology* **18**, e3000600 (2020).

69. Martin, S. G. & Arkowitz, R. A. Cell polarization in budding and fission yeasts. *FEMS microbiology reviews* **38**, 228–253 (2014).
70. Van Hooff, J. J., Tromer, E., van Dam, T. J., Kops, G. J. & Snel, B. Inferring the evolutionary history of your favorite protein: a guide for molecular biologists. *BioEssays* **41**, 1900006 (2019).
71. Russell, J. J. *et al.* Non-model model organisms. *BMC biology* **15**, 1–31 (2017).
72. Johnson, B. R. & Lam, S. K. Self-organization, natural selection, and evolution: Cellular hardware and genetic software. *Bioscience* **60**, 879–885 (2010).
73. Wedlich-Soldner, R., Wai, S. C., Schmidt, T. & Li, R. Robust cell polarity is a dynamic state established by coupling transport and GTPase signaling. *The Journal of cell biology* **166**, 889–900 (2004).
74. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology* **24**, 79–88 (2006).
75. Khmelinskii, A. *et al.* Incomplete proteasomal degradation of green fluorescent proteins in the context of tandem fluorescent protein timers. *Molecular biology of the cell* **27**, 360–370 (2016).
76. Daalman, W. K.-G., Sweep, E. & Laan, L. A tractable physical model for the yeast polarity predicts epistasis and fitness. *Philosophical Transactions of the Royal Society B* **378**, 20220044 (2023).
77. Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications (1970).
78. Bodnar, O. & Elster, C. On the adjustment of inconsistent data using the Birge ratio. *Metrologia* **51**, 516 (2014).
79. Christiano, R., Nagaraj, N., Fröhlich, F. & Walther, T. C. Global proteome turnover analyses of the yeasts *S. cerevisiae* and *S. pombe*. *Cell reports* **9**, 1959–1965 (2014).
80. Carmona-Gutierrez, D. *et al.* Guidelines and recommendations on yeast cell death nomenclature. *Microbial Cell* **5**, 4 (2018).
81. Kass, R. E. Raftery, AE (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773 (6).
82. Cone, A. *Yeast DAPI Staining V1* 2016. <https://www.protocols.io/view/Yeast-DAPI-Staining-eigbcbw> (2023).
83. Munoz, A. J., Wanichthanarak, K., Meza, E. & Petranovic, D. Systems Biology of Yeast Cell Death. *FEMS Yeast Research* **12**, 249–265. ISSN: 15671356. <https://academic.oup.com/femsyr/article-lookup/doi/10.1111/j.1567-1364.2011.00781.x> (2012).

An approach to extract genetic interactions from saturated transposition analysis in yeast.

People can only be free if they are truly educated

Jose Marti

Abstract

Genetic interactions serve to understand how genes influence biological processes and phenotypes. Genetic interactions occur when the fitness effect of one gene mutation is modified by the presence of mutations in other genes. The budding yeast, *S. cerevisiae*, is a key model for studying genetic interactions due to its well-characterized genome. However, systematic high-throughput techniques for computing genetic interactions in yeast are limited. Saturated Transposition Analysis in Yeast (SATAY) offers a promising solution. While SATAY has identified gene functions and essentiality changes, its potential for generating high-throughput fitness landscapes to compute genetic interactions remains underutilized. In this study, we introduce a fitness model transforming SATAY read-outs into relative fitness values, addressing biases in relating SATAY read-outs with fitness. Genetic interactions are derived using a multiplicative fitness model that accounts for fitness variances across replicates. The fitness model and resulting genetic interactions provide a foundation for guiding further investigations using low-throughput experiments.

Parts of this chapter have been published in Biorxiv [1].

3.1. Introduction

In the post-genomic era, a crucial task is to attribute functions to genes and unravel the organization of gene networks governing cellular processes. The budding yeast *S. cerevisiae* has stood as a pivotal eukaryotic model in tackling this challenge. An invaluable asset in this endeavor has been a compendium of strains wherein each yeast coding sequence is systematically deleted. This compendium has enabled the evaluation of each gene's impact on any measurable phenotype. These genomic methodologies can be expanded to methodically quantify genetic interactions.

A genetic interaction between a pair of genes is observed when individual genetic mutations combine, resulting in an unexpected phenotype distinct from the phenotypes caused by the individual mutations [2]. Genetic interactions arise from the intricate and complex network of interconnections between an organism's genes. Understanding how genes interact provides insights into their shared roles to fulfill a specific function. For instance, when a double knockout of a gene pair leads to a non-viable phenotype, contrary to the viability of individual phenotypes, it suggests that both genes function in redundant or compensatory pathways. Systematically studying such genetic interactions, known as synthetic lethal interactions, can reveal potential drug targets, particularly for tumor removal in cancer treatments[3, 4].

Specifically, in the case of pairwise genetic interactions, we identify that genes i and j interact if the fitness of the double mutant differs from the multiplication of the single mutant fitnesses, denoted as $f_{ij} \neq f_i f_j$. Here, f_{ij} represents the fitness of the double mutant of genes i and j , while f_i and f_j represent the fitness values of the respective single gene mutants. Expressing these fitness values in normalized units is crucial to avoid unit mismatches after their multiplication. To define the strength and type of the genetic interaction, we quantify $\epsilon_{ij} = f_{ij} - f_i f_j$.

There are two major classes to classify digenic, that is, between a pair of genes, genetic interactions: Positive and Negative interactions. One special case within the negative interaction category is synthetic lethality. These classes follow from the value of ϵ_{ij} . **Positive interaction** means that $\epsilon_{ij} > 0$, **Negative interaction** refers to $\epsilon_{ij} < 0$, and **Synthetic lethality** is when $f_{ij} \sim 0$, or $\epsilon_{ij} \approx -f_i f_j$.

Various techniques, particularly high-throughput methods, have been developed to systematically study genetic interactions across different species. One prominent approach is Synthetic Genetic Arrays (SGA) [5], which involves assessing the fitness (based on population colony size) of a vast yeast deletion library comprising single, double, and, more recently, triple mutants[2]. Through quantifying the fitness effects of millions of yeast mutants and grouping genes with similar interaction profiles, researchers have established what is known as the global yeast genetic interaction network [6]. However, using this technique to investigate how genetic interactions change with genetic backgrounds or under different experimental conditions is intractable because of the necessity of generating millions of yeast knockouts to explore the effects of nearly all genes makes this approach extremely challenging to implement in standard wet labs.

Since the detection of the fitness of multiple mutants in different experimental and genetic environments is central to determining genetic interactions, newer and more flexible techniques for high-throughput fitness measurements are emerging, which are based on the simultaneous measurement of the contribution of large genetic variants to the fitness of cells or organisms. This type of assay is called pooled fitness assays [7, 8]. This procedure is then followed by next-generation sequencing to detect changes in mutant frequencies [7–9]. One example of these techniques is Barcode sequencing (Bar-seq), where mutant genomes are tagged with a unique nucleotide sequence that allows their identification after sequencing [10]. Although the pooled

assay of Bar-seq greatly facilitates fitness assessment, it has the same issues as SGA regarding the laborious steps required for mutant library construction [11]. In addition, the number of mutants that can be assessed in a single Bar-seq experiment is relatively low (typically between 1,000-5,000) due to the limited availability of unique barcodes [11–13].

An alternative to Bar-seq for creating mutant libraries is transposon insertion sequencing (TIS). TIS methods use transposons to randomly move within DNA molecules, creating a library of mutants by disrupting genes. Unlike Bar-seq, TIS doesn't rely on having specific barcodes available for each mutant [14–16]. Typically, transposon mutagenesis is efficient enough to produce libraries with more than 100,000 mutants, and a single library will often contain multiple mutants carrying disruptions at different locations in the same gene [15, 17, 18]. As mutants are identified based on the transposon insertion site by sequencing the transposon-genome junction, the genomic library preparation and bioinformatics analysis of TIS data are relatively complex.

Recently, a TIS method named SATurated Transposon Analysis in Yeast (SATAY) has been developed for the yeast *Saccharomyces cerevisiae*. This procedure was implemented and optimized by the Kornman Lab [17]. The detailed experimental protocol can be found in section 3.4.

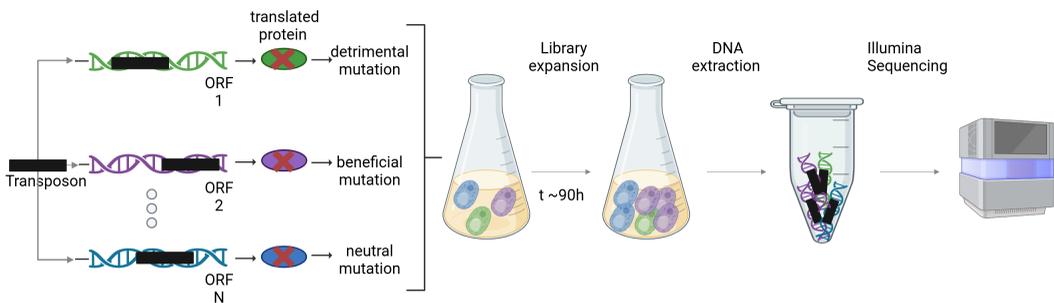


Figure 3.1. Basis of SATAY. The transposon element (black rectangle) is induced to jump to random genomic locations. If the transposon hits an open reading frame (ORF), it will cause the inability to translate a fully functional protein. Once all cells from the population carry a transposon insertion, they will be grown for about 90 hours to amplify their fitness differences. Next, the DNA is extracted from a part of the population and sequenced.

The SATAY procedure first triggers the translocation of a transposon to a random genomic location, see fig. 3.1. Suppose the transposon is inserted in an open reading frame (ORF), in that case, it will hinder the correct translation of that coding sequence to a fully functional protein, rendering a mutation that is effectively a complete gene knockout. Although this is what is generally assumed, there are some cases in which the transposon insertion affects the translation of one non-essential protein domain, rendering a functional truncated protein, [17, 19]. In a population of cells, every cell will have only one independent transposition event. The next part of the procedure amplifies the differences in fitness effects each mutation will confer to the population. The mutations can be detrimental or essential, beneficial or neutral. The first one will render very few or no copies of this mutant in the population. The second will trigger an excess of this mutation in the population, and the last one will be as the population average, see fig. 3.2C. This step is called library expansion and lasts approximately 90 hours. It is effectively a competition assay where all possible genetic mutations are mixed. Lastly, one part of the expanded population

3. An approach to extract genetic interactions from saturated transposon analysis in yeast.

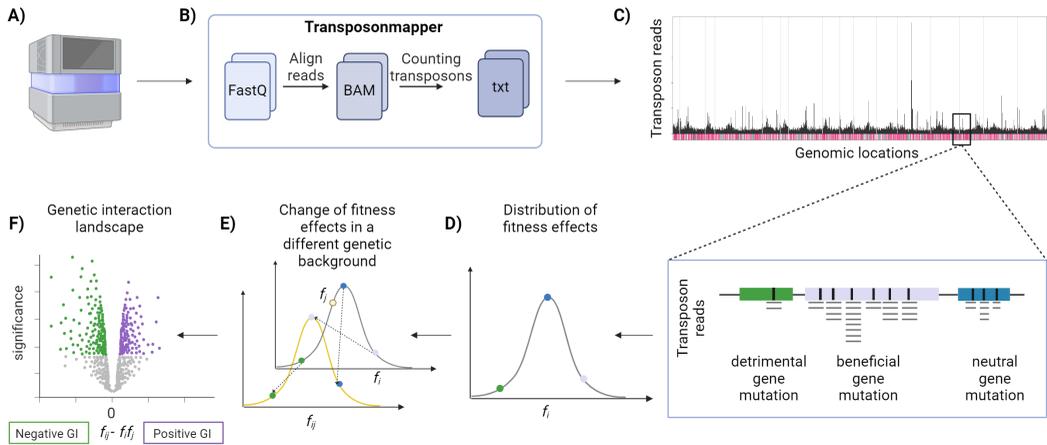


Figure 3.2. From transposon sequencing to genetic interaction landscapes. **A)** Following the sequencing phase, **B)** Transposonmapper is employed to align the sequencing data with the reference genome. This process involves mapping the genomic locations and quantifying the abundance of each transposon across the genome. The outcome of this procedure facilitates **C)** the identification of transposon insertions, their respective sequencing reads over genomic regions, and subsequently, the derivation of an **D)** inferred distribution of fitness effects (f_i). **E)** By repeating this process in different genetic backgrounds, we can establish genetic interaction profiles. **F)** Visualization of these interactions is achieved through volcano plots, where the y-axis represents the significance value (e.g., p-value), and the x-axis denotes the genetic interaction score, defined as $f_{ij} - f_i f_j$. In these plots, negative and positive genetic interactions (GI) for a specific gene are represented by green and purple dots, respectively

is used for DNA extraction and further sequencing of the transposon junctions to the genome fig. 3.2A.

The analysis of the sequencing data shows where each transposon landed in the genome and how many copies were found fig. 3.2B. The number of transposons linked to the exact location is proportional to the number of cells carrying that specific mutation. This parameter is called the reads of a particular transposon insertion, fig. 3.2C, which is vital to computing the fitness value associated with that transposon. Furthermore, by analyzing the fitness effects of a specific gene, fig. 3.2D, in different pertinent genetic backgrounds, fig. 3.2E, it is possible to compute the genetic interactions at a genome-wide scale of the corresponding gene, fig. 3.2F. For instance, to quantify the genetic interaction of *geneX*, we should generate SATAY libraries in wild-type (to obtain the single gene knockout fitness) and full $\Delta geneX$ (to get the double gene knockout fitness) knockout backgrounds.

Traditionally, SATAY has been primarily used to identify essential (e.g., its gene knockout leads to cell death) and toxic (e.g., its native expression is detrimental to the population growth rate) genes by studying their impact on cell growth under specific environmental and genetic conditions. The working principle of SATAY allows us to distinguish between essential and toxic genes based on the effects of transposon insertions in coding sequences. If a transposon disrupts an essential gene, then that gene will harbor few or no transposons. Conversely, if a transposon disrupts a toxic gene, it leads to multiple copies of such mutation and therefore abundant transposons in that gene, see fig. 3.2C.

In the original publication [17], SATAY was used to identify changes in the fitness contribution of genes in different genetic and environmental contexts through a one-on-one comparison of their transposon insertion density. While practical, this approach can only be used to identify fitness changes of the same gene across datasets but not of different genes within the same dataset. In addition, it neglects possible subtle differences reflected in the read count but not in the insertion density.

Here, we describe an approach to generate quantitative fitness maps from read count data obtained from a single SATAY dataset. We propose to incorporate the read counts associated with each insertion as the proxy for the fitness value of that specific transposon insertion. We follow an exponential model to estimate the fitness values during the growth of the mutant library from the average read counts per gene from each library. Our definition of fitness relies on a population growth model, equivalent to the growth rate of a specific mutant relative to the *wild type* value. We also include extra factors influencing the complex relationship between the read counts and fitness values. We show that contrary to the expectation, including domain spatial information in the resultant fitness value per gene did not improve the predicting power of using fitness values from SATAY to determine essential genes. However, this result could be attributed to our limited database knowledge to identify the protein domains.

Furthermore, although our method demonstrates reasonable reproducibility across technical and biological replicates, the fitness distribution it yields notably diverges from those obtained using other techniques. Nevertheless, we argue that despite these differences, SATAY-derived fitness values remain valuable for detecting genetic interactions. This is because SATAY captures a broad range of fitness levels, unlike other methods that are primarily centered around wild-type fitness levels.

3.2. Results

Fitness from the read counts

To determine the fitness effects genome-wide in a specific genetic background, we need to estimate the growth rate (μ_g) of each yeast mutant (g) based on the read counts ($r_{i,g}$) that map to the transposon insertion locations i of gene g .

We adopt a Malthusian model[20] to represent the fitness, as it assumes that the population growth rate is proportional to the size of the population at any time. In our case, we neglect the competitive aspect of the SATAY fitness assay due to the lack of information on the population at different time points. Consequently, we assume that all mutants are in the exponential growth phase, and we disregard any potential overcrowding effects or limited resources that could influence the population growth rate. Using this Malthusian model, we aim to estimate the growth rate for each yeast mutant to further identify genetic interactions across the entire yeast genome in the specific genetic background of interest.

It follows from the Malthusian model that the growth of the population of cells carrying a mutation g , P_g , will behave as eq. (3.1).

$$\frac{dP_g}{dt} = \mu_g P_g, \quad \mu_g > 0 \quad (3.1)$$

where μ_g is the growth rate at which the population carrying a mutation in gene g grows. After

integrating both sides:

$$P_g = P_{0g} 2^{\mu_g t}, \text{ where } P_{0g} \text{ is the initial population carrying the mutation on gene } g \quad (3.2)$$

$$\mu_g = \frac{\log_2 \frac{P_g}{P_{0g}}}{t} \quad (3.3)$$

The population of cells carrying a mutation in gene g is represented by the number of reads per insertion along gene g . To estimate each mutant's growth rate (μ_g), we consider the average value of reads per insertion along the gene as its representative value.

At the start of the library expansion step, we assume that there is only one representative of each mutant. In other words, we assume there is no cell division at the beginning of the library expansion step, so the initial population size (P_{0g}) equals 1. The growth rate values regarding the read counts per gene are expressed in eq. (3.4).

$$\mu_g = \frac{\log_2(\frac{1}{N} \sum_{i=0}^N R_{ig})}{t} \quad (3.4)$$

Where R_{ig} are the reads counts for the insertion location i in gene g , and N is the number of insertions along same gene g . The primary underlying assumption is that most mutations confer a neutral fitness effect to the organism[21, 22].

$$\mu_{ref} = \frac{\log_2(\text{median}(\vec{R}_{ig}))}{t} \quad \forall i, g \quad (3.5)$$

We use the median instead of the mean since we seek to avoid the effects of the outliers in our reference. Finally, the normalized growth rate for the population of cells carrying a mutation in gene g looks like eq. (3.6).

$$\bar{\mu}_g = \frac{\log_2(\frac{1}{N} \sum_{i=0}^N R_{ig})}{\text{median}(\log_2(\vec{R}_{ig}))} \quad (3.6)$$

By applying expression eq. (3.6) to all genes, we will generate the distribution of fitness effects of each gene inactivation for the studied particular genetic background.

Correcting biases for the fitness quantification

Correcting for insertions near the edges of the genes

Transposon sequencing studies often assume transposon insertions to result in complete loss of gene function [15]. Under this assumption, read counts obtained from different insertion sites within the same gene can be considered replicate measurements of the same gene deletion mutant, and their averaging is justified. However, it has been found that insertions near the 5' and 3' end of a gene are less likely to result in a complete loss of gene function [23]. Similarly, studies using SATAY have reported that genes considered to be essential can sometimes tolerate insertions close to the gene ends while central regions remain empty [17]. If this higher tolerance to insertions near gene ends is a general phenomenon affecting all genes, averaging the read counts of the entire coding region of a gene would create a bias in the fitness estimate.

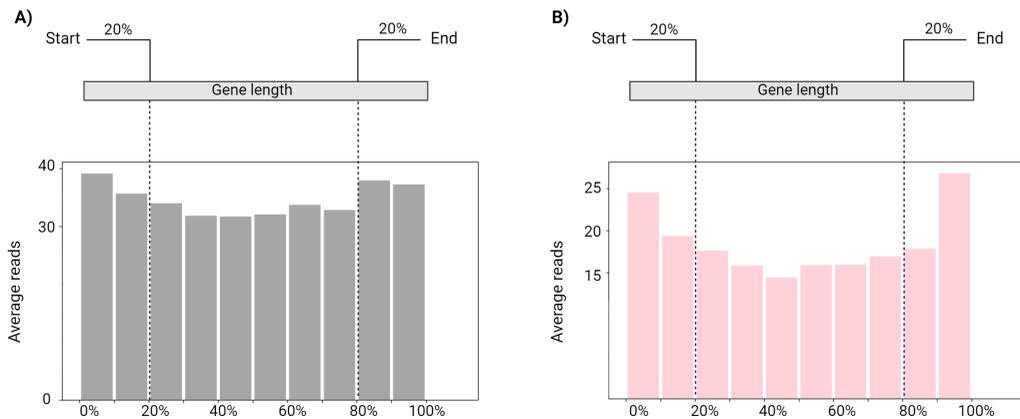


Figure 3.3. The average reads enrichment bias is illustrated at both the beginning and end of each gene. Genes are categorized into ten equal regions based on their length, as shown on the x-axis. The y-axis depicts the mean reads for each segment across all genes. **A)** Represents all genes. **B)** Focuses specifically on annotated essential genes.

Thus, we check if we could observe a general trend of transposon insertions located near the gene edges at the genome-wide scale. To do so, we segmented the open reading frames of all annotated genes in the genome into ten equally sized bins, such that each container amounts to 10% of the coding sequence. We then calculated the average number of reads per transposon insertion mapped to the respective segment for each segment. For genes annotated as non-essential, we observed that insertions within the first and last 10% of a gene tend to acquire more reads than insertions in the central 80% of the gene, although this effect is weak, see A) from fig. 3.3. However, this non-uniformity was more pronounced for genes annotated as essential by [24], see B) from fig. 3.3. This difference between essential and non-essential genes is expected, as the fitness difference between a complete and partial knockout should typically be larger for essential genes.

Interestingly, the read count profile of essential genes shows that the bias towards higher read counts is stronger for insertions close to the stop codon. This likely reflects the mechanism for gene inactivation of the MiniDS transposon, which is based on creating gene truncations by introducing several early stop codons in the open reading frame [17]. Thus, insertions that lead to truncations close to the C-terminal part of the protein will often still allow the protein to (partially) retain its functionality.

Since insertions that do not lead to complete gene knockouts invalidate the averaging over the read counts of different insertion sites, we decided to exclude all insertions that map within the first or last 10% of a coding region for fitness calculations.

Correcting for the preferential insertion of MiniDS at pericentromeric sites

The proposed fitness model using the logarithm of the average of read counts over the central 80% of the gene length assumes that each read count retrieved by sequencing correlates truthfully with the growth rate of that specific mutant. However, we may neglect one crucial aspect related to rare mutants with poor growth and high death rates. These mutants will likely be lost during the sampling steps that lead up to sequencing. Hence, the insertion sites associated with

these mutants remain undiscovered in the final SATAY dataset; see fig. 3.4A.

While the loss of rare transposon mutants forms the basis for identifying essential genes, it can distort the relation between mutant fitness and the read counts per insertion site. In this case, we assume that rare mutants harbor insertion sites with zero reads that cannot be detected after sequencing. The non-inclusion of these sites will artificially overestimate the fitness values of rare mutants because the average is taken over fewer insertions than over the total including the undiscovered ones.

3

One approach to determine the number of insertion sites with zero reads for each gene is to use the global insertion density to infer an expected insertion rate. To that end, we use the known bias of MiniDs transposons to insert preferentially in regions close to its excision site. According to the specific experimental design of SATAY, transposons are more likely to insert in the pericentromeric areas rather than in more distal regions. To verify the existence of this centromere bias in our insertion data, we plotted the cumulative insertion count as a function of varying distances to the centromere, see fig. 3.4B. We observed an enrichment of transposons in genomic regions closer to approximately 200 kb from a centromere. This effect did not result from a lower density of essential genes in the pericentromeric areas. If left unaddressed, this bias would lead to underestimating the expected insertion rate in pericentromeric regions and overestimating the anticipated number of transposons in low-growth mutants in those regions.

We assessed a polynomial model to correct for the centromeric bias, fig. 3.4C. The same method was followed to model the changes in the insertion rate up to distances 200 kb from the centromere while assuming a constant insertion rate beyond 200 kb. This assumption is based on the idea that the centromeric bias is a ‘memoryless’ feature such that the insertion rate no longer depends on the distance to the centromere beyond a certain point.

By differentiating the polynomial fit concerning distance, we obtain a smoothed estimate (trended) of the insertion rate changes that reflect its global characteristics but do not contain the local fluctuations caused by the specific properties of the genes that reside there. Overall, the insertion rate varied between approximately 0.07 bp^{-1} for regions close to the centromere to 0.03 bp^{-1} for more distal areas. Because centromere bias should not play a significant role for regions very far from the centromere, we artificially flattened the fitted curve by setting a constant rate for positions located further than 200 kb away from the centromere. In summary, we use the following equation to determine a bias-corrected insertion rate λ :

$$\lambda_{r_c} = \begin{cases} a_0 + a_1 r_c + a_2 r_c^2 + a_3 r_c^3 & \text{if } r_c < 2 \cdot 10^5 \\ \lambda(r_c = 2 \cdot 10^5) & \text{if } r_c \geq 2 \cdot 10^5 \end{cases} \quad (3.7)$$

r_c is the distance from the centromere in basepair and a_{0-3} are the coefficients obtained from least squares polynomial fit. Because we assume that the insertion rate remains approximately constant throughout a gene’s coding sequence, the expected number of insertions for a gene g is calculated by multiplying the insertion rate with gene size:

$$E(X_g) = \lambda(r_{c-g})L_g \quad (3.8)$$

With r_{c-g} , the distance from gene g to the centromere measured from the gene’s center, and L_g the length of the gene in base pairs. The number of zero read count sites is then estimated to be equal to the difference between $E(X_g)$ and the number of observed insertions $O(X_g)$ when

$E(X_g) > O(X_g)$, see A) from fig. 3.4. When the number of observed insertions exceeds their expected value, we conclude that no unobserved insertion sites exist for that gene.

To determine whether our model of the expected insertion rate was effective at identifying the larger fraction of undiscovered insertion sites in annotated essential genes, we plotted the difference in the expected and observed insertion counts for all genes, see E) from fig. 3.4. Here, we consider this difference to reflect the number of undiscovered transposon insertions after sequencing. The plot showed that the expected number of insertions for nearly all essential genes is higher than those found from the read data. For annotated non-essential genes, the difference between expected and observed insertion counts follows a Gaussian-like distribution around zero that partially overlaps with the distribution for essential genes.

Thus, including undiscovered insertion sites provides additional information on essentiality, although insufficient to distinguish annotated essential from non-essential genes. However, our results show that including these zero-read count sites to estimate fitness is crucial to resolving differences in the fitness distribution's lower end.

For the fitness calculation, it specifically means that we first compute the expected insertion rate of every gene using eq. (3.7) depending on the distance of this gene to the centromere. Then, we compare the expected insertion, from eq. (3.8), with the actual number of insertions in the sequencing data. If the expected insertions are larger than the observed ones, we add zeros to the insertion counts. Otherwise, we keep the actual sequencing data. Lastly, we apply eq. (3.6) for the final fitness value of each gene [1].

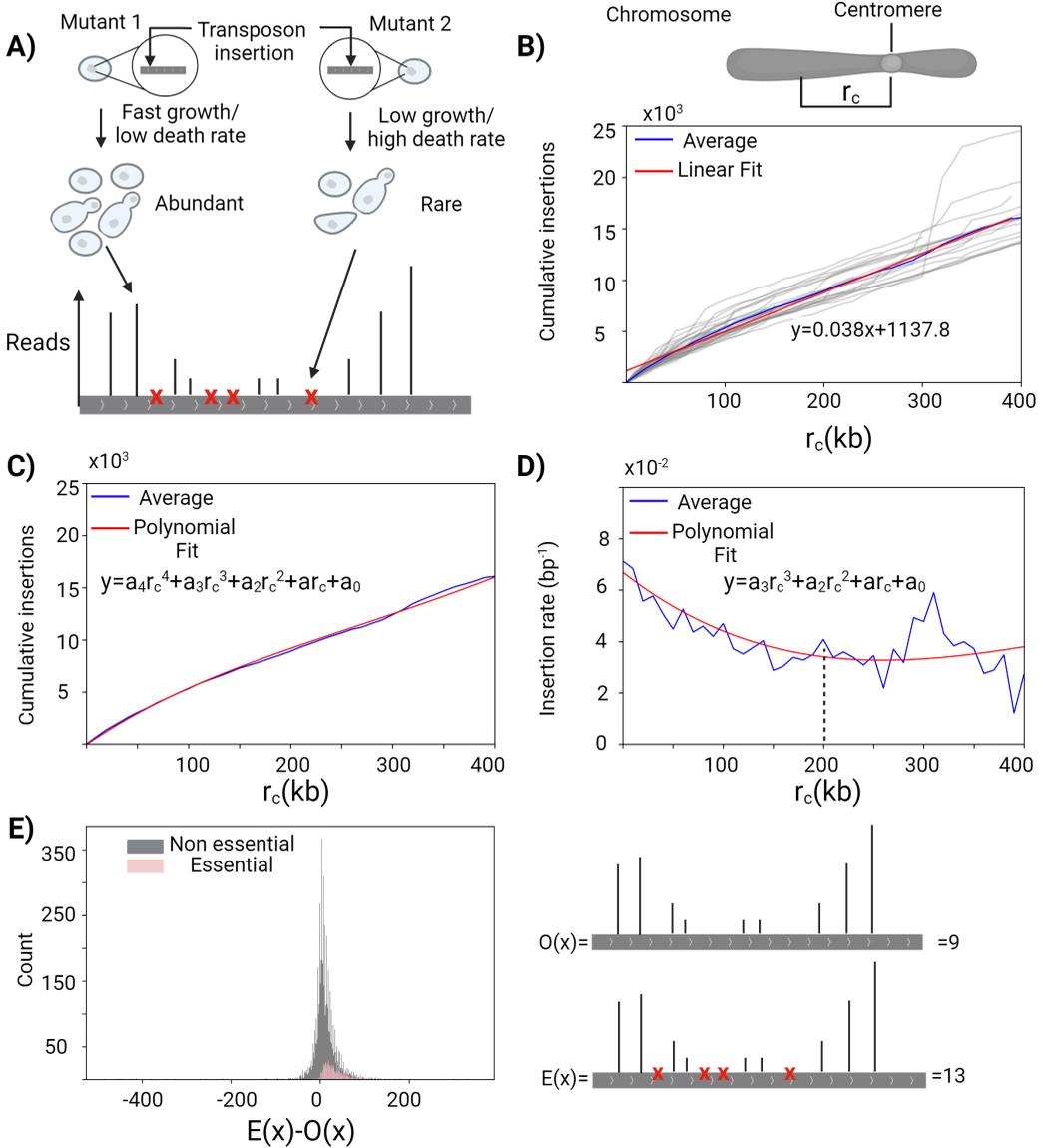


Figure 3.4. Correcting for centromere bias to determine genes' expected transposon insertion rate. **A)** The goal of calculating the expected insertion rate is to estimate the number of insertion sites that produce no reads because the mutants were lost during the population sampling due to low abundance. **B)** The empirical insertion rate depends on the distance of a gene to the centromere. To visualize this bias, we determined the number of transposons that mapped within a distance r_c from the chromosome centromere for different values of r_c , as done previously by [17]. The plots of the cumulative insertions for the two halves of each chromosome are shown as grey lines. The averaged curve is shown as a blue line. The non-zero intercept of a linear fit of the portion of the average curve for distances $r_c < 400kb$ demonstrates the centromere bias in our dataset. **C)** The averaged cumulative plot for distances $r_c < 400kb$ was fitted with an exponential function: $-1.5 \cdot 10^{-19}r_c^4 + 2.7 \cdot 10^{-13}r_c^3 - 1.5 \cdot 10^{-7}r_c^2 + 7.8 \cdot 10^{-2}r_c - 124$ and a 4th-order polynomial. **D)** The approximated empirical insertion rate is fitted by a polynomial function: $-6 \cdot 10^{-19}r_c^3 + 8 \cdot 10^{-13}r_c^2 - 3 \cdot 10^7r_c + 0.07$. After distances of $r_c > 200kb$, we consider the insertion rate a constant function. **E)** The difference between the expected ($E(X)$) and observed ($O(X)$) transposon insertions for essential and non-essential genes. While the average difference is close to zero for non-essential genes, it becomes positive for essential genes.

3.2.1. Fitness refinement using spatial information

Averaging the reads along a gene neglects the positional effects of different transposon insertions on fitness. This property leads to possible distortion of the fitness of essential genes or genes that contain toxic domains for specific genetic and environmental conditions.

For instance, if a transposon hits a sequence that codes for an essential protein domain responsible for cell survival (e.g., transcription), the resulting truncated protein may strongly impact the cell's fitness. On the other hand, if transposon insertion occurs outside this essential domain, the truncated protein may retain its essential function, leading to a milder fitness effect. When examining essential genes from [24], we observed regions void of transposons within these genes. These genomic regions likely encode essential protein domains responsible for the critical function of the protein, as also observed by [17]. Similarly, areas that tolerate transposon insertions would correspond to truncated versions of the essential protein that are dispensable for the cell's viability.

To address this issue, we consider individual transposon insertions and their specific locations within the gene to include the contribution of their positional effects. This approach aims to capture the fitness impact of individual domains within genes.

$$\mu_g^* = \max(|\mu_g - \mu_i|) \quad \forall i \quad (3.9)$$

We specifically utilized the annotated protein domains from the Pfam database [25], which is part of the extensive resource Interpro [26]. Pfam comprises a comprehensive collection of multiple sequence alignments and hidden Markov models encompassing numerous common protein domains. One limitation of this approach is that we restrict our domain search to the ones annotated in that database. Thus, we don't account for the genes that do not have any domain annotated in Pfam. We primarily used this method to explore if considering the fitness of the gene as the fitness of the domain with the most substantial effect (for those who have annotated domains from Pfam) will benefit the classification of genes as essential (or genes required for growth in our conditions).

To incorporate these protein domains into our analysis, we obtained all the relevant domains from the Pfam database, see fig. 3.5A and transformed their protein coordinates into genomic coordinates (disregarding genes with introns). We generated a vector of fitness values for each gene, corresponding to the number of annotated domains, by employing the same procedure described in eq. (3.6) and accounting for the insertional bias to pericentromeric regions for each domain.

To ensure that fitness values come from sufficiently transposon-covered genes, we excluded genes with poor transposon density without transposons localized at the annotated domains on that gene. As a hard cut-off, we impose that genes that harbor less than five insertions are discarded for the fitness computation if the transposons localized in their annotated domains are less than two. Genes that have domains with a scarce number of transposons (less than two insertions) are kept if the expected number of transposons over the gene is larger than five, see these cases in fig. 3.5B.

We compute the fitness using eq. (3.6) for each domain within every gene. Then, we select the fitness value from the domain with the strongest effect compared to the average fitness of the entire gene, as described in eq. (3.9). For the gene products that contain a single domain, we

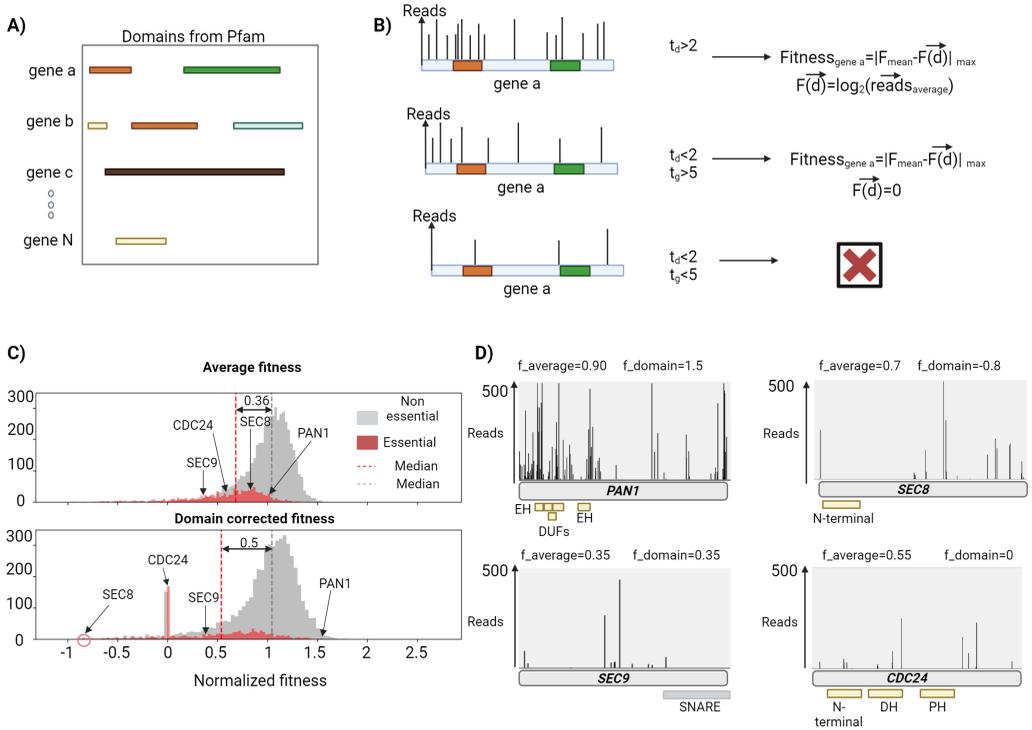


Figure 3.5. Including positional effects of transposons by including protein domain information from Pfam database. **A)** We use the Pfam domain library annotation to extract all domain coordinates from every gene product. **B)** Three cases of insertion profiles over domains are differentiated for fitness estimation. Domains with more than two expected insertion ($t_d > 2$), receive a fitness score as the logarithm of the average reads over the domain. The domains that harbor less than two transposons have fitness zero, if only if the gene in which the domain resides has more than five transposons ($t_d < 2$ and $t_g > 5$). The domains with less than two transposons on genes with less than five insertions are discarded for further analysis ($t_d < 2$ and $t_g < 5$). **C)** DFE for the average fitness and the domain corrected fitness. In gray and red are the non-essential and essential fitness distribution. The median of the essential DFE is shifted to smaller values when using the domain correction, while the median of non-essential remains almost unchanged compared to the average method. **D)** Examples of essential genes that exhibit different properties and how the different fitness methods capture them.

take the smaller fitness value between the average of the whole gene and the fitness from the single domain as the representative fitness of the gene knockout.

As expected, introducing spatial information on the fitness calculation will amplify the differences between genes since we no longer consider that all transposons over the gene will cause the same effect. This is evidenced when comparing the distribution of fitness effects (DFE) with and without the domain correction, that we appreciate a wider distribution for the fitness values coming from the domain with the strongest effect (standard deviation $\sigma = 0.91$ vs. $\sigma = 0.31$). We detect that the effect of disrupting domains can recover many genes whose knockout may have a strong deleterious effect on the cell. Specifically, 582 genes with fitness below or equal to zero from the domain correction versus 161 genes from the average fitness. Genes that show a fitness value above the reference value of 1 are similarly retrieved by both methods (3181(domain correction) vs. 3197(average)).

To explore the differentiation of essential genes in the DFE, we examine their fitness values (see fig. 3.5D). Interestingly, 303 essential genes display zero or negative fitness values, compared to 104 when domain correction is not applied. Notably, essential genes exhibit higher fitness values with domain correction, possibly due to transposon-enriched regions within them. For instance, the gene *PAN1* shows a fitness value of 1.5 with domain correction, compared to 0.9 when considering only the average. This gene contains multiple annotated domains, including EH (Eps15 homology) domains associated with endocytosis and vesicle transport, as well as several domains of unknown function (DUFs) near its N-terminal [27].

There were 84 genes out of 1117 annotated essentials from [28] that did not have Pfam domain annotations. Those genes did not get the correction. Hence, their fitness values will be the averaging over their reads, like the gene *SEC9*. Interestingly, this gene possesses a domain annotated from another database (Panther [29]) in the C-terminal called SNARE ("SNAP REceptors"), which is the domain that confers the primary function of this essential protein. The transposon insertion profile of this gene has the region at the end of the gene void of transposons, which directly leads to the essentiality of this domain, as also [17] revealed.

The case of the gene *SEC8*, with few insertions in the N-terminal but enough insertions elsewhere in the gene, receives a negative fitness, and it lays at the negative extreme of the DFE using the domain correction due to the addition of zero reads insertions alongside its domain region. Thus, the average over the domain region is less than 1. Yet, if all the reads over the gene are averaged, the normalized fitness increases to around 0.63, around the median value of all essential genes using this method.

CDC24 is another example of an essential gene that tolerates insertions in dispensable regions. Yet, it receives a zero fitness when applying the domain correction because its pleckstrin homology domain is void of transposons. In contrast, if averaging all reads over the genome, this gene knockout receives a considerably high fitness value of 0.56.

In conclusion, introducing spatial information on the fitness calculation seems to amplify the fitness difference between annotated essential and non essential genes. For the case of essential genes, this method allows to distinguish domains that may seem toxic in our conditions, thus displaying a high fitness value compared to the average. Yet, it is not yet clear the mechanism by which the MiniDs transposon insertions will lead to specific protein truncations, especially for the transposons inserted in the 5' of the gene, that lead to the production of the C-terminus part of the protein like the case of the essential gene *SEC9*. This is surprising since several stop codons in all frames of the transposon should interrupt translation and prevent the production of the essential C-terminus. [17] also observed this phenomenon, and they speculate that the production of the C-terminus is enabled by spurious transcription events.

However, it is still uncertain whether this approach can benefit the discernibility between annotated essential and non-essential genes based on fitness values.

3.2.2. The domain corrected fitness does not improve the performance to predict existing essential genes

To quantitatively classify the performance of each fitness method to predict existing essential genes, we transformed the problem into a binary classification task to assess how the fitness values obtained from the protein domain correction can be used to predict gene essentiality. We categorized genes as essential or non-essential based on the median fitness values of each DFE for essential genes. This resulted in fitness thresholds of 0.68 and 0.53 for the average and

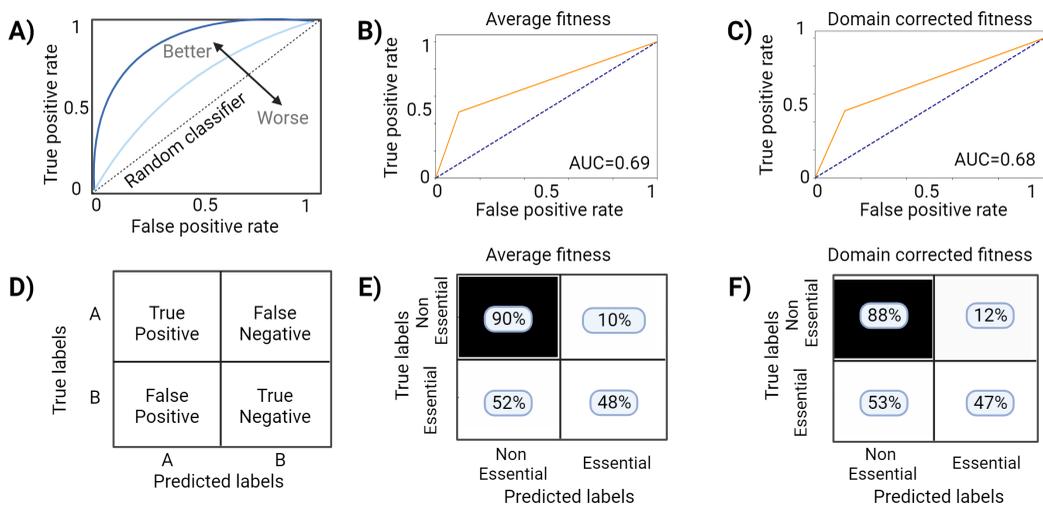


Figure 3.6. The fitness refinement based on protein domains does not improve essentiality prediction. **A)** Sketch exemplifying the meaning of the ROC curve. The x-axis is the false positive rate, and the y-axis represents the true positive rate. The farthest from the diagonal line, in the y-axis direction, the better the classifier. **B)** ROC curve for taking average fitness values as probabilities to classify a gene as essential. The area under the curve (AUC) indicates how accurate this classifier is, and in this case, is 0.69, which means that for every 100 genes analyzed, 69 are properly classified in either category. **C)** Same as **B)** but with the corrected fitness values as a classifier. The area under the curve is 0.68. **D)** Sketch of the confusion matrix. This matrix is a 2x2 matrix that represents the accuracy of specific properties to predict a variable with two cases: A or B. The diagonal represents the accuracy of predicting either A (True positive) or B (True negative). Off-diagonal positions represent different misclassifications, corresponding to false positive or false negative ratios. **E)** Confusion matrix for the average fitness as a classifier. **F)** Same as **E)** but with the domain corrected fitness as classifier.

domain-corrected fitness models, respectively.

We compared our predictions to the annotated essential genes from the [28] database. To evaluate the accuracy of our predictions, we utilized Receiver Operating Characteristic (ROC) curves, which are graphical plots used to visualize the diagnostic ability of binary classifiers. The area under the ROC curve (AUC) measures the classifier's performance, with a higher AUC indicating better predictive power. The ROC curve for the average fitness model showed an AUC of 0.69, indicating a moderate accuracy in predicting gene essentiality. Likewise, the ROC curve for the corrected fitness model showed an AUC of 0.68, which renders no change in the predicting power of this method compared to the other.

To understand the specifics of our gene essentiality predictions, we analyzed the confusion matrix shown in D) of fig. 3.6. When using the average fitness as a classifier of essentiality, 48% of the actual essential genes, according to [24], were correctly identified. On the other hand, non-essential genes were predicted with high accuracy, achieving a 90% correct classification.

When considering the corrected fitness as a classifier of essentiality, we retrieved 47% of the referenced essential genes that were correctly identified as essential (according to [24]). We also found that 12% of the non-essential genes were predicted as essential by the corrected fitness model. These genes could potentially be essential genes specific to our experimental settings. Lastly, 88% of the non-essential genes were classified correctly as non-essential.

In summary, using the corrected fitness values did not improve the prediction of essential genes compared to the average fitness. There are several reasons why the domains annotated in Pfam for annotated essential genes may not be responsible for its essential function. Firstly, essential genes may have other domains not annotated in Pfam, which play a more crucial role in their essential function. Secondly, essential genes may also have domains that can become toxic for the cell without another domain in the protein. In such cases, disrupting these domains will benefit the population's fitness. For instance, the DUF domain from *PAN1* and the DH domain from *CDC24* work together with the PH domain. The PH domain is invariably located immediately C-terminal to the DH domain, and this invariant topography suggests a functional interdependence between these two structural modules[30]. Biochemical data have established the role of the conserved DH domain in Rho GTPase interaction and activation and the role of the tandem PH domain in intracellular targeting or regulation of DH domain function[30]. And thirdly, the annotated essential genes database may be not the ground truth for our specific experimental and genetic background conditions.

Despite the lack of benefit from applying the protein domain corrected fitness in predicting gene essentiality, it is still an important tool to predict critical domains for function and, thus, subtle fitness effects from protein truncations.

3.2.3. The domain corrected fitness moderately benefits the reproducibility among experiments

One crucial aspect of the applicability of SATAY for global fitness maps of gene disruption mutants is that the fitness estimates are reproducible between replicates of the same genetic background. To that end, we compared the fitness values between two technical replicates obtained from the wild-type strain. These replicates were split after DNA extraction before proceeding with the transposon junction PCR amplification step (as described in section 3.4). Thus, those replicates are subjected to two-fold noise sources: the noise from the PCR amplification and the one from the sequencing process, see A) from fig. 3.7. We assessed the degree of agreement between the independent measurements by plotting the fitness estimates of each gene disruption using the average and the domain-corrected fitness models.

When examining the individual fitness per gene from both models, high variance is evident, especially at low fitness values. From a biological perspective, we can interpret that mutants with low read counts are likely sick mutants struggling to survive in the population. Consequently, their fitness values highly depend on the specific micro-environment they experienced during the experiment, particularly during the library expansion step. This dependence on micro-environmental conditions contributes to these mutants' observed variance in fitness values.

Furthermore, another aspect that will also contribute to reduce the noise among replicates is the selection of replicates after the PCR amplification and thus only affected by the sequencing noise. Posterior studies, like [1], have used this type of replicates with a similar fitness approach based on the average reads counts per insertion site from the 80% central part of the gene, and their fitness reliability increases up to 97%.

Despite the mentioned noise sources, the domain-corrected fitness model shows more reproducibility across technical and biological replicates, see C) and E) from fig. 3.7. Especially for the case of the biological replicate, the increase is moderate, according to the person correlation coefficient among the fitness values between the replicates, which goes from 0.76 to 0.83. The fact that the domain corrected fitness takes an average over a smaller genomic region than

3. An approach to extract genetic interactions from saturated transposition analysis in yeast.

the whole gene artificially reduces the noise over the reads related to the sequencing process. Moreover, many regions of the gene are not considered in many cases because there are no annotations in the Pfam database in those regions.

In summary, our analysis confirmed that considering the protein domain with the strongest effect results in higher reproducibility, across experiments, of fitness values for each gene knock-out compared to considering the average read counts over the entire gene. In the next section, we describe how we use the generation of fitness effect of gene disruptions for all genes to compute genetic interactions.

3

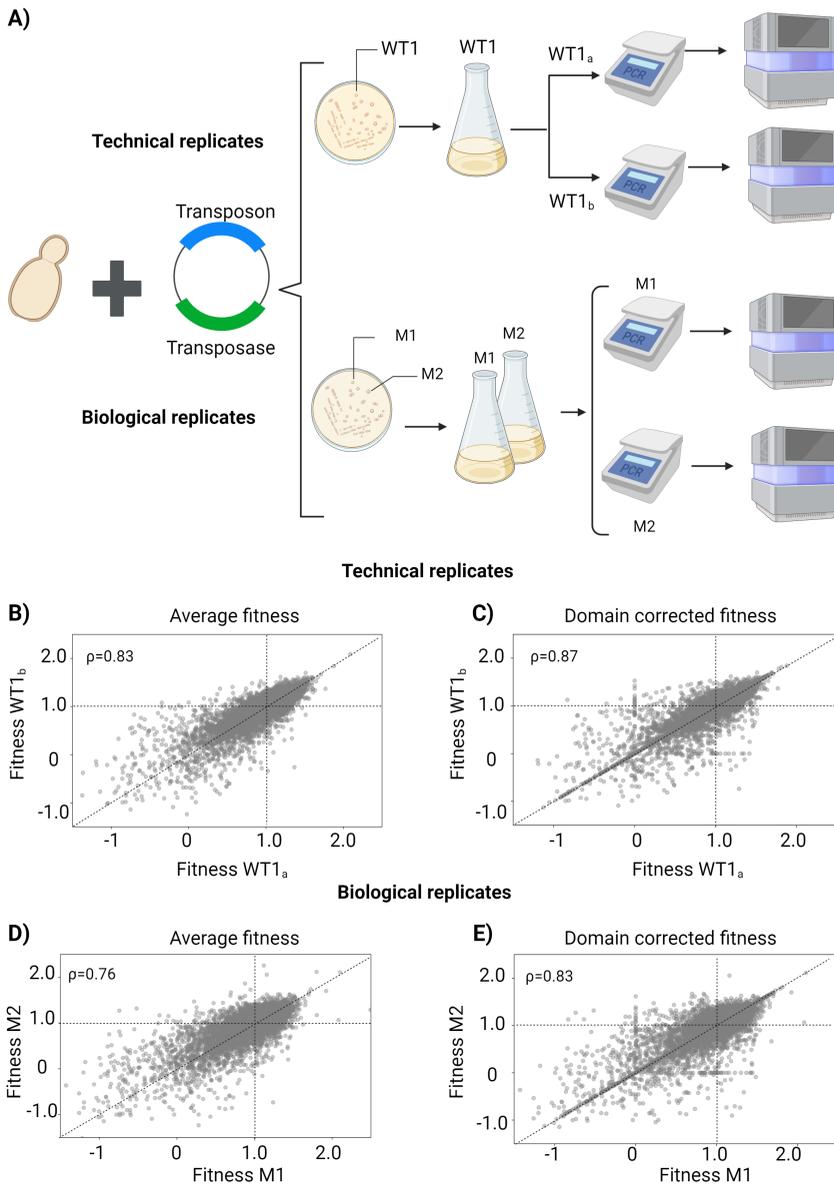


Figure 3.7. The domain corrected fitness improves the reliability among technical and biological replicates. **A)** The technical replicates come from the same transformed yeast cell (WT1) with the plasmid carrying the MiniDs/Transposase and their subsequent mutant library. Before PCR, the extracted DNA was split into two batches (WT1_a and WT1_b), which we defined as the technical replicates. Biological replicates come from two transformed yeast cells (M1 and M2) with the plasmid carrying the MiniDs/Transposase. **B)** The estimated fitness effect of gene disruptions for all genes of technical replicate WT1_b plotted against its estimated value in technical replicate WT1_a, using the average model. The identity line (diagonal dashed line) is a reference for perfect correlation between the two replicates. The Pearson correlation coefficient is shown at the top left of the graph. **C)** The estimated fitness effect of gene disruptions for all genes of technical replicate WT1_a plotted against its estimated value in technical replicate WT1_b, using the domain-corrected fitness model. The identity line (diagonal dashed line) is a reference for perfect correlation between the two replicates. The Pearson correlation coefficient is shown at the top left of the graph. **D)** Same as **B)** but for the biological replicates M1 and M2. **E)** Same as **C)** but for the biological replicates M1 and M2.

3.2.4. Identification of genetic interactions

Traditional methods that map genetic interaction among multiple genes of interest rely on constructing the relevant yeast mutants, which carries the limitations of possible secondary undesirable mutations during construction. This work proposes to use SATAY, which generates yeast mutants *de novo* identified by transposon sequencing, to map genetic interactions using the fitness of gene knockouts from the read counts contained in sequencing data. One challenge with this approach is the lack of a unique reference system for all the fitness estimates in each genetic background. The fitness values are normalized within the same library but not across libraries; thus, they can not be directly compared without making extra assumptions. One way to tackle this limitation is by centering the DFE in a particular mutant background in the fitness value that this specific mutant had in the wild-type library. This assumes that most gene knockouts in the background mutant do not change its original fitness. A graphical explanation of this additional normalization is shown in A) from fig. 3.8.

To obtain the genetic interaction scores from the relevant DFE, we follow the multiplicative model, shown in eq. (3.10), which is graphically illustrated in B) from fig. 3.8.

$$\epsilon_{ij} = f_{ij} - f_i f_j \quad (3.10)$$

One of the challenges with this approach is that the DFEs of the same mutant across different experiments are subjected to variability that is unfeasible to control. The inherent noise sources from the population sampling, PCR, the sequencing process, and the researcher's experimental practice reduce the robustness across the fitness estimates. To deal with this inherent noise in all the samples, we compute a T-test over the DFEs in the mutant background against the product of the individual gene knockouts in the wild-type background. This will lead to excluding genes with such a high variability among replicates that their fitness values are unreliable in computing their genetic interaction, see C) in fig. 3.8.

Lastly, we assemble our genetic interaction scores with their corresponding p-values into a volcano plot (D) fig. 3.8) highlighting the significant positive and negative genetic interactions of each target gene.

For the computation of genetic interactions, we employ the fitness estimates from the average read count over the gene because despite its increased variability across experiments in the same genetic background in contrast with the domain corrected fitness. This is due to the systematic artificial exclusion of data points that the domain correction imposes on the fitness estimate. Given that the Pfam database is limited and that we lack a proper data-driven method to extract domains from our datasets, we decided not to continue with this method for the search of genetic interactions. Next, we investigate how well the fitness estimates from SATAY of gene knockouts align with data from other high-throughput experimental techniques that rely on the *Saccharomyces* Genome Deletion Project[24, 28].

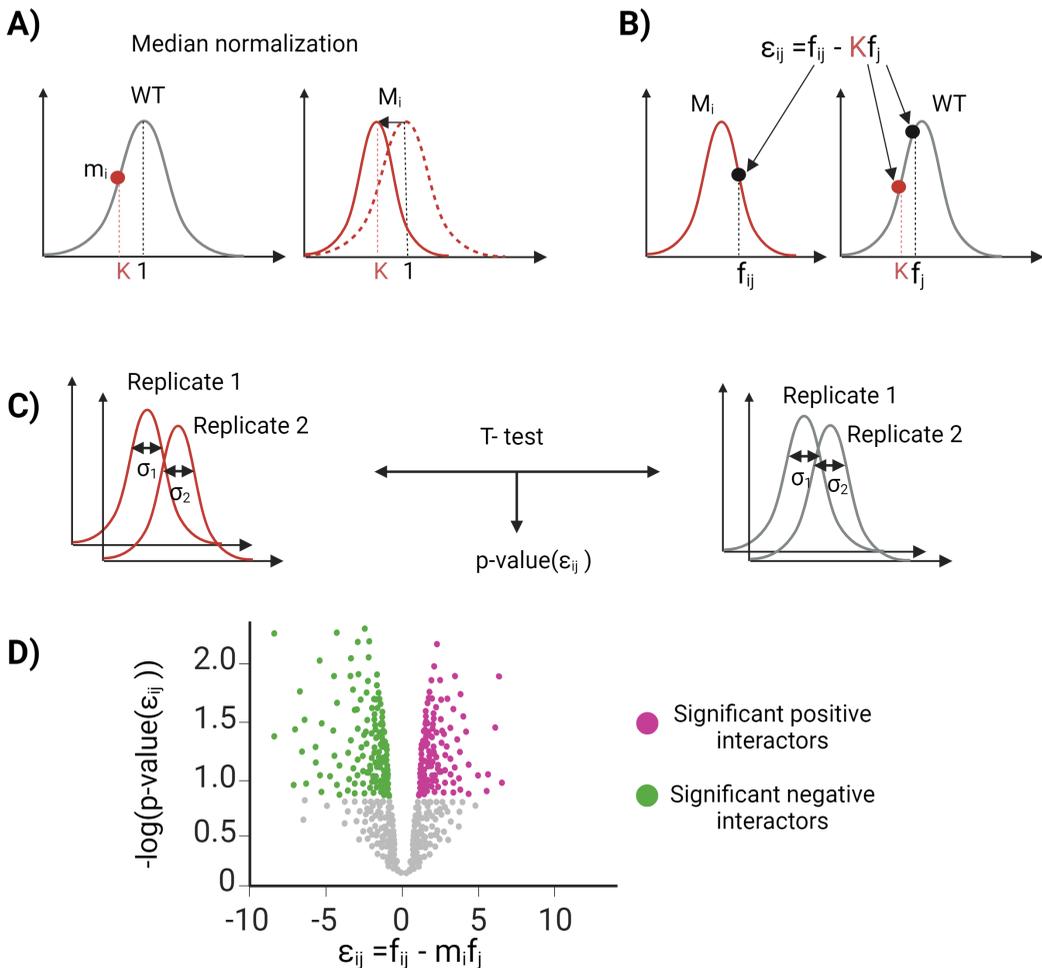


Figure 3.8. Genetic interaction methodology using fitness from SATAY transposon sequencing data. **A)** Additional normalization to the DFE in a mutant background (M_i) to ensure that the relative fitness values are centered around 1. Specifically, the median normalized DFE is multiplied by the gene knockout (K) fitness in the wild-type (WT) background. **B)** The genetic interaction scores (ϵ_{ij}) are obtained from a multiplicative model ($f_{ij} - K f_j$). Where f_{ij} is the fitness of gene j in M_i background, and f_j is the fitness of gene j in WT. **C)** We incorporate the variability in variances of the DFEs across different replicates by computing a T-test that compares the DFEs from the wild-type background against the DFEs from the mutant background. **D)** The identification of a specific gene's positive and negative genetic interactions is through representing the genetic interaction scores against their p-values.

3.2.5. Fitness estimates from SATAY correlate poorly with reported estimates based on the Yeast Gene Deletion collection

We selected two high throughput techniques that rely on the *Saccharomyces* Genome Deletion Project [24, 28] to provide fitness estimates of gene knockouts, namely Synthetic Genetic Arrays (SGA) and competitive fitness measurements using Bar-seq Illumina sequencing [31].

These techniques are all high-throughput methods that allow for the rapid computation of fitness values of a pre-made yeast knockout library. However, they differ in their approaches to

computing the fitness of single gene deletions.

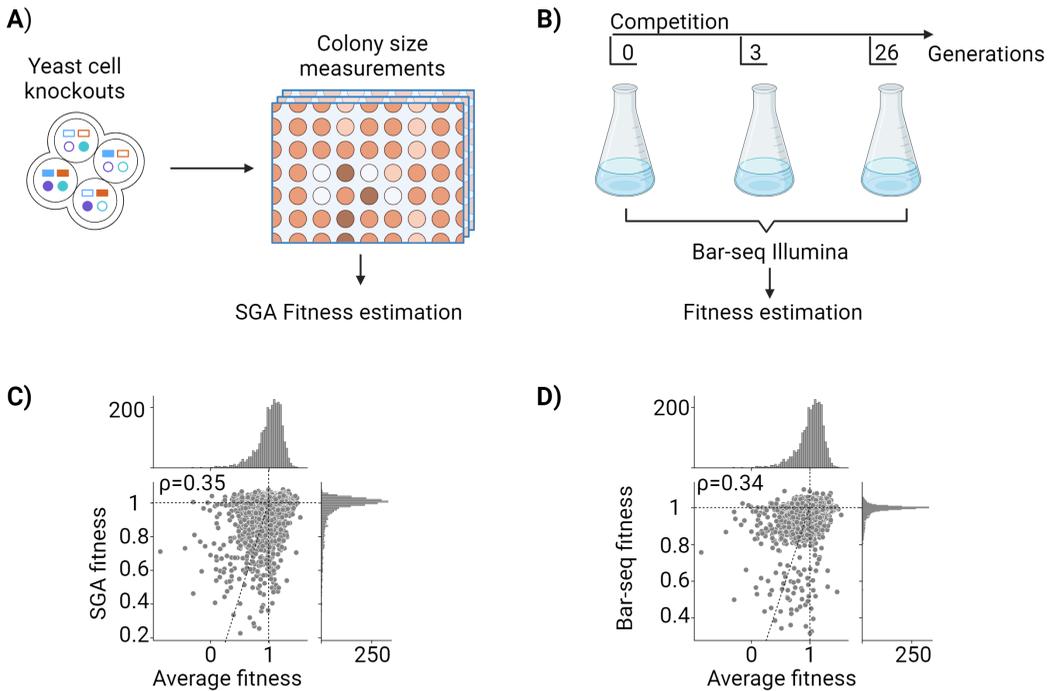


Figure 3.9. Poor positive correlation between fitness estimates from SATAY and SGA and Bar-seq fitness estimates. **A)** SGA performs colony size measurements of yeast knockout libraries to determine the fitness of single gene deletion mutants [32]. **B)** A competition experiment that uses Bar-seq to determine the relative number of mutants with respect to a reference in different time points as a proxy for the fitness of single gene deletion mutants. [31]. **C)** Fitness estimates from the average model from SATAY correlate poorly with fitness estimates from SGA experiments. **D)** Fitness estimates from the average model from SATAY correlate poorly with fitness estimates from the competition assay based on Bar-seq.

With SGA, the fitness is determined by measuring the colony size of a yeast mutant population, see A) from fig. 3.9. This technique is especially important since almost all genetic interaction existing data is from this procedure, which we employ to compare the genetic interactions from SATAY.

The case of Bar-seq experiments is particularly interesting as they used read counts from sequencing of unique barcodes attached to every strain to determine the relative abundance of the knockout with a wild-type strain as a proxy of their fitness. All the tested yeast knockouts are grown together, and their relative abundance is measured in three time points, see B) fig. 3.9. The measurements are taken up to generation 26, which may be sufficient time for sick mutants to acquire beneficial mutations that artificially shift their fitness towards higher values.

The average fitness estimates from SATAY correlate poorly with SGA fitness estimates of single gene deletions, C) fig. 3.9. Firstly, the SATAY DFE ranges over more values than its counterpart. Thus, many genes that are considered to have near-neutral effects once deleted using SGA will be deleterious for the SATAY fitness estimates. And the contrary is also valid. Many mutants with deleterious effects by SGA are classified as neutral for SATAY. We computed the Pearson correlation coefficient, which quantified the correlation between two variables assuming a linear model

[33], among the fitness values of the same gene. For the case of SGA and SATAY, the coefficient is 0.35, which indicates a poor positive correlation, C)fig. 3.9.

In the case of Bar-seq experiments, the correlation is also poor among the two methods. As expected, the Bar-seq DFE distribution is even narrower than for the case of SGA around neutral fitness values, which can be explained by its dependency on the competition setup, where small differences in growth are harder to assess. In contrast to SATAY, which also relies on a competition assay, the barcode dependency limits the number of mutants the experiment can hold.

Although the correlation values are below 0.5, they are positive, and it suggests that genes with high/low fitness in one technique also tend to have higher/lower fitness values in SATAY. This indicates a consistent trend in the fitness behavior of genes across different techniques, even though the absolute fitness values do not match closely.

One resemblance between the SGA and Bar-seq experiments is their low variance around a fitness value of 1. This could be a result of the fact that in pre-constructed gene deletion collections, mutants usually pass through several rounds of replication after their construction before they are subjected to a fitness assay. This allows secondary mutations that can potentially mask the defects caused by the primary gene deletion to set in the population. It has indeed been shown that several mutants in the yeast gene collection harbor such compensatory mutations[34, 35]. Libraries created with SATAY are less likely to suffer from secondary mutations because the fitness assay usually follows directly after library creation. As a consequence, fitness values based on measurements of the yeast gene deletion collection may overestimate the number of neutral genes compared to results obtained with SATAY, which is in agreement with our observation that the DFE obtained by [32] and [31] is much narrower than ours.

The second reason for the disparity observed in the fitness values from SATAY compared to [32] and [31] could be attributed to the genetic backgrounds of the yeast mutants used in each experiment. Whereas we used W303, [32] and [31] used S288C.

Moreover, it is shown that SATAY has the ability to resolve smaller differences in growth between mutants compared to other methods. This capability to detect subtle fitness differences can be particularly useful in identifying genetic interactions and uncovering the roles of specific genes and protein domains in various cellular processes. However, the differences with the mentioned techniques indicate that the fitness values obtained through the analysis of SATAY datasets may be condition-specific and cannot be generalized without further considerations of experimental design.

3.3. Discussion

In this chapter, we have shown a method to quantify the fitness of gene deletion mutants based on data from the transposon mutagenesis screen SATAY. This method attempts to deal with the non-trivial relation between read counts and fitness.

Transposon insertions near gene edges often yield higher read counts, affecting the observed fitness of gene disruptions. To address this bias, we exclude insertions close to gene edges. However, transposon fitness effects also vary depending on the specific protein domain affected. Grouping insertions by annotated Pfam domains helps capture these nuanced effects but may overlook insertions outside annotated domains, reducing resolution. Interestingly, while the domain approach does not improve essential gene classification compared to average read counts, it offers greater robustness across replicates, likely due to noise reduction from averaging over

smaller regions. This may mitigate stochastic sequencing errors associated with longer coding sequences.

Regarding the disparities in predicting essential genes, it is key to acknowledge that caution is required when comparing the results of our study with the essential genes obtained by the [28] study. Differences in growth environments and genetic strain backgrounds used in our study and theirs may influence gene essentiality outcomes. Recent studies have highlighted that gene essentiality, even within the same species, is not always conserved when altering the genetic background and environment. There are two classes of essential genes identified: core essential genes that are universally required for viability and conditional essential genes that vary in essentiality under different genetic and environmental conditions [36–38]. A comprehensive understanding of gene essentiality must consider these variations in different contexts.

In the study by [28], essential genes were determined by individual heterozygous gene deletions followed by viability tests of haploid segregants. The results showed that approximately 18% of yeast genes are essential for viability under nutrient-rich media conditions using the laboratory strain S288C. However, it was also found that among all known essential genes in S288C, about 17% are dispensable essential genes [39], meaning that they can be bypassed by suppressor mutations. These conditional essentials were enriched for genes with paralogs and genes encoding membrane proteins while being depleted for genes encoding protein complexes. The interactions among functionally related genes frequently drove bypass suppression interactions.

Additionally, [40] demonstrated the existence of conditional essentiality across two closely related yeast strain backgrounds, Σ 1278b and S288C, which differ in approximately 1% of their genomes. In their study, they identified 75 proteins present in Σ 1278b but not in S288C, and about 1% of the genes in S288C were specific to this background. The observed conditional essentiality was attributed to a complex set of background-specific modifiers that influence the phenotype of mutations and contribute to differences in essentiality between individuals.

We utilized the strain background W303 [41], which shares more than 85% of its genome with S288C but differs at over 8000 nucleotide positions, leading to changes in the sequences of 799 proteins [42], representing approximately 15% of the W303 genome. Considering the comparison made with S288C and Σ 1278b, we expect to have at least 15% of essential genes unique to W303. Moreover, it is necessary to note that the growth context in the study [28] differed from our experimental conditions. They relied on nutrient-rich media, while we used synthetic media with adenine dropout to expand our library. As described in [43], environmental factors can influence gene essentiality, and some essential genes may become dispensable under non-standard growth conditions. [28] found that around 2% of genes can be specific to the external environment.

Considering all these factors, it is reasonable to expect at least 30% differences in gene essentiality due to our specific experimental procedure. Specifically, we found that 48% of our predictions correctly match the standard dataset from [24], and 10% of the essential genes were unique to our experimental procedure. These results highlight the significant effects of specific genetic backgrounds and environmental contexts on gene essentiality.

Furthermore, we propose a method based on the fitness estimates from SATAY to compute statistically significant genetic interactions. Despite the given the variability in fitness values among technical and biological replicates in our samples, we claim that by using this method, we can explore a larger mutant space to find significant and novel genetic interactions among specific genes of interest.

One of the challenges in using SATAY is the lack of a real reference for the fitness values, which can be especially problematic when studying different genetic backgrounds or environmental conditions. The fitness values obtained from SATAY are based on the assumption that most transposon insertions in the 80% central part of the gene have neutral effects on the cell. While this assumption may hold for the wild-type genetic background, it may not be valid for other genetic backgrounds or under different environmental conditions. To address this issue, it is crucial to identify more suitable references that allow for proper fitness comparisons across different conditions. This could involve using other experimental methods to measure gene fitness in specific conditions and using these results as a reference to validate and calibrate the SATAY fitness values. Combining SATAY with complementary assays and employing appropriate references can enhance the accuracy and reliability of fitness assessments for gene disruptions and essentiality in budding yeast under diverse genetic and environmental contexts.

In conclusion, applying SATAY on different genetic backgrounds in budding yeast can provide insights into the fitness implications of disrupting various genes or detecting gene essentiality for specific conditions. However, it is required to complement SATAY results with more specific assays, such as microscopy or population growth studies of target mutant yeast strains, in the same laboratory conditions where SATAY was performed. These additional experiments will help to carefully test and refine the fitness values given by SATAY.

3.3.1. Acknowledgements

Thanks to Dr. Enzo Kingma to facilitate the scripts to compute the global insertion profile based on the distance of each insertion to the centromere. Thanks to Prof. Benoit Kornmann and Dr. Agnes Michel for sending the plasmid carrying the MiniDs/Transposase system and sequencing the SATAY mutant libraries from the ylic133 and ylic136 strains, as well for their help on troubleshooting our SATAY experiments through the SATAY online forum.

3.4. Materials and Methods

SATAY procedure

Plasmid transformation. The transposon system used is a modified non-autonomous Maize transposon (MiniDs) which interrupts the ADE2 gene. A Galactose-induced hyperactive transposase (Ac) mobilizes the transposon. ([44],[45]). Homology-directed repair of ADE2 then confers adenine prototrophy to the transposed cells. A screen begins by transforming a plasmid (pBK549) encoding the Galactose-induced transposase, the ADE2 gene interrupted by the MiniDs transposon, and a URA3 selection marker into an *ade2- ura3-* strain [19]. The plasmid pBK548 is a centromeric plasmid¹. See a map of it in fig. 3.10

We use plasmid pBK549 to transform our yeast cells deficient on *ade2-* and *ura3-*. Details of this procedure can be found in section 3.4.

Sanity check. Once the cells have been transformed with the pBK549 plasmid, it is crucial to test the ability of the cells to repair the ADE2 gene by spontaneous homologous recombination². Hence, the so-called sanity check consists of plating the same colony from a transformation plate

¹These are considered low copy vectors and incorporate part of an ARS along with part of a centromere sequence (CEN). These vectors replicate as though they are small independent chromosomes and are thus typically found as a single copy.

²Homologous recombination is a type of genetic recombination in which genetic information is exchanged between two similar or identical molecules of double-stranded or single-stranded nucleic acids (usually DNA as in cellular organisms but maybe also RNA in viruses).[46]

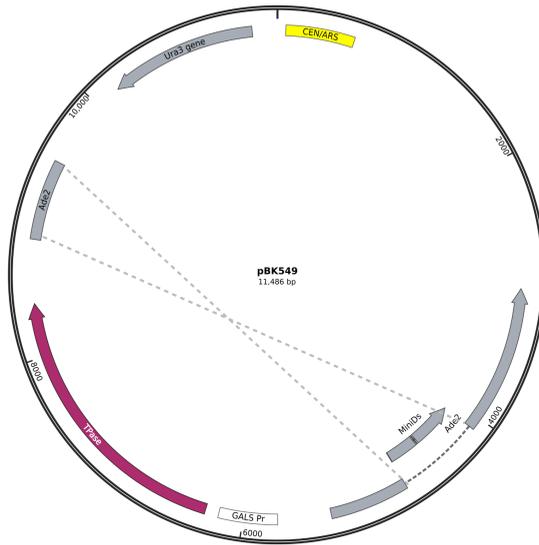


Figure 3.10. Map of the plasmid (pBK549) to transform into yeast cells to perform a SATAY screen. Its main features are the active transposase in red with the upstream Galactose promoter. The ADE2 gene is interrupted by the MiniDS transposon and its repair fragment (dotted lines), the URA3 gene, and CEN/ARS element.

in solid growth media lacking uracil and adenine. The goal is to select the colonies that show full growth in uracil deficient media and few growths in adenine deficient media. See a graphical representation of this in fig. 3.11, left panel.

It is also possible to have full growth in adenine deficient media. However, despite its clear proof that the cells can spontaneously repair the ADE2 gene, during the SATAY screen, we would have many cells producing adenine but not due to transposition but to homologous recombination, which will translate into a high background of false positives in our assay. Thus, we want to discard also those clones that show this behavior. If, on the contrary, there is no growth in adenine deficient media, it will imply that the repair mechanism does not work on those cells. A reason for it could be that the initial population of cells that undergone the plasmid transformation had a mutation in a different gene like ADE1 from the adenine biosynthesis pathway³ thereby hindering the production of the adenine aminoacid, even though the ADE2 gets repaired. Another possibility is that they have already recombined out ADE2, so there will be no transposition. These cases are graphically shown in fig. 3.11, right panel. We continue with the clones that show from 2 to 20 colonies in adenine deficient growth media.

Preculture. Once we have selected the clones that passed the sanity check, we are ready to start the SATAY procedure. The first step in the screen is called preculture, where isogenic yeast cells

³<https://www.phys.ksu.edu/gene/genefaq.html>

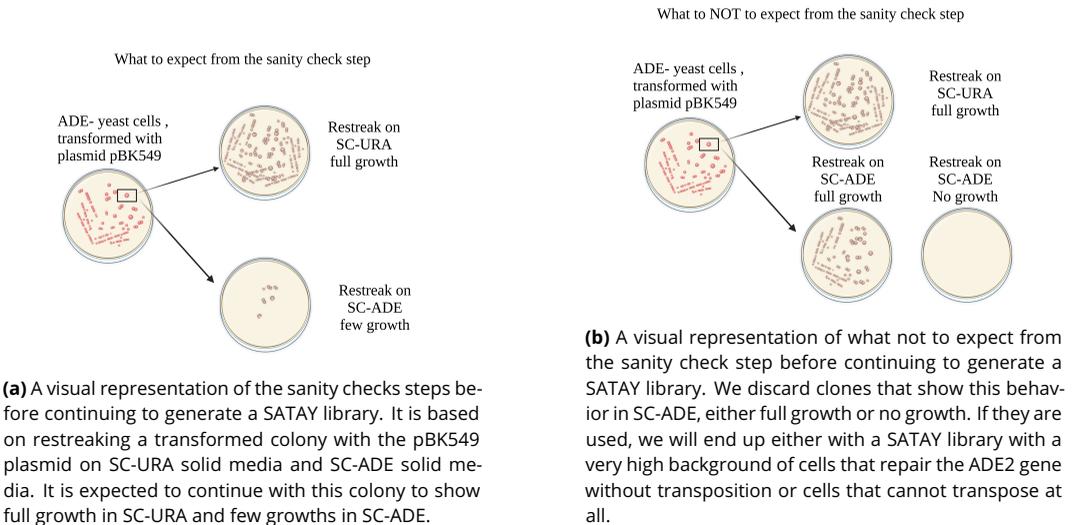


Figure 3.11. What to expect and not expect from the sanity check step after plasmid transformation.

are grown in SD-URA+0.2% glucose +2% raffinose during 20 hours. This step will amplify the cells with the plasmid without any transposase expression. The typical volume to use per clone is from 20-30mL, and the incubation is at 30°C.

Induction. This is the second step of the protocol and the most important one. In this step, the cells from the preculture are diluted into SD-URA + 2% galactose. The galactose will trigger the galactose promoter upstream of the transposase, thereby expressing it. Once this transposase is active, it will excise the transposon in the plasmid. Subsequently, the transposon will jump into the yeast genome and insert in a random location, with 60% of probability [19]. And since it is a low copy plasmid, it will produce one transposon so that we won't have the situation of more than one transposon jumping into different places of the genome of one cell. The induction is performed in 100-200mL per clone, and it takes approximately 50-52 hours. This step is where we make sure the genome gets saturated with transposons, and it will set the resolution of our assay, that is, the number of transposons per bp of the genome. Our type of transposon does not have any target sequence of insertion. However, it tends to insert in intergenic regions outside coding regions and insert in nearby excision sites locations. In our case, since it comes from a centromeric plasmid, it will have higher chances to insert in pericentromeric⁴ regions of the chromosome [48]. Notably, this step monitors the number of cells that become ADE⁺ during this time. This is done by plating in three time points (0,24h,48h) 200ul of cells in SC-ADE + 2% glucose, counting the number of cells visible in the plate and then extrapolating to 1mL. The measurement at t=0h will give us the background that we start with for the induction; these are cells that have already repaired the ADE2 gene without transposition. As a rule of thumb, we should expect 200-400 ADE⁺ cells per mL.

Reseed. After the transposition induction in a population of yeast cells, we end up with transposons in hopefully all locations from the yeast genome reflected in the places it jumped to for

⁴pericentromeric (comparative more pericentromeric, superlative most pericentromeric) (genetics) Situated near, or on each side of, the centromere of a chromosome [47]

every cell in the population. Now, it is time for the reseed, which has two main functions: pausing the jump of transposons to the genome by stopping the activation of the galactose promoter and allowing the expansion of the population size and thereby the differences in fitness effects of the different transposon insertions. From this step, we get after the sequencing the reads per transposon insertion, which biologically means how many cells harbor that transposon at that location. We assume that all the mutations start with one copy in the population for the reseed. Hence all the growth should happen in this step. The volume per clone in this step is 3L of SD-ADE +2% galactose. The incubation is at 30°C for about 80 hours. Details of the SATAY procedure in yLIC133 and yLIC136 strains are reported in <https://leilaicruz.github.io/Experimental-jupyterBOOK/journal/2020-08/2020-08-10-SATAY-bem2d-nrp1d-WT.html>

DNA extraction. After the reseed, the harvest of the DNA from each clone follows. Importantly in SATAY, we cant keep the mutants generated for every library. Instead, what is analyzed is the DNA of the population by deep sequencing procedures where it is mapped the site of insertion of every transposon in the genome. Therefore, we first need to obtain the DNA from each library. For that, we follow the protocol [17]. An overview of all the steps towards sequencing is represented in fig. 3.12. The details of the DNA extraction for the yLIC133 and yLIC136 samples is described in <https://leilaicruz.github.io/Experimental-jupyterBOOK/journal/2020-08/2020-08-18-DNA-Prep-SATAY.html>

DNA sequencing. Sequencing involves the following steps⁵: Digestion of genomic DNA with two four-cutter restriction enzymes, ligation by Ligase-mediated circularization of the DNA, PCR of the transposon-genome junctions using outward-facing primers, Illumina-sequencing of the combined PCR products.

Digestion. This step is performed with the restriction enzymes: DpnII and NlaIII. The goal is to chop the DNA into pieces. These enzymes have cutting sites inside the MiniDs. Hence we will create hybrid DNA fragments of genomic DNA with part of the transposon.

Ligation. In this step, the DNA with a transposon gets circularized by using a ligase enzyme.

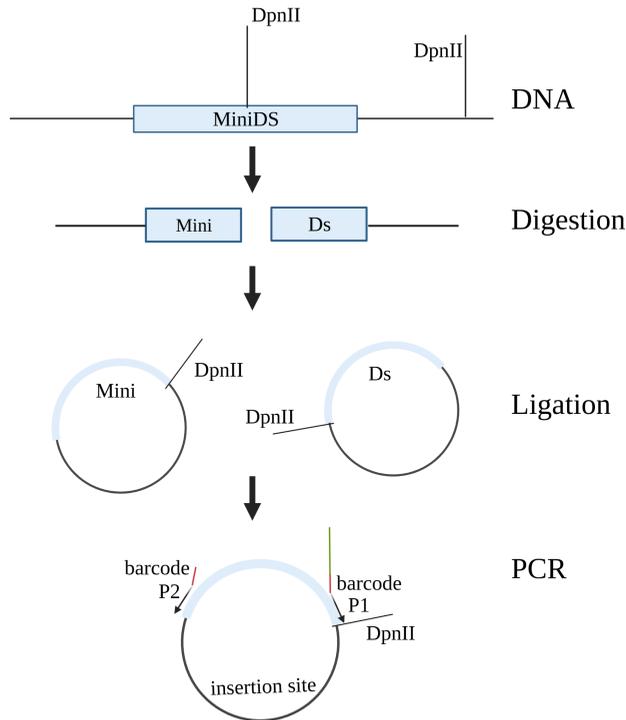
PCR. A specific Polymerase Chain Reaction (PCR) protocol is performed on the ligated fragments to amplify the library for sequencing. Per library, we do two PCR rounds on the DpnII digested-circularized pool and in the NlaIII digested-circularized pool. We use barcodes, which are 8bp DNA sequences, used to differentiate between DpnII and NlaIII pools after the sequencing. For the DpnII pool we use primer 1_DpnII and primer 2_DpnII from table 3.3. For the NlaIII pool we use primer 3_NlaIII and primer 4_NlaIII from table 3.3. To design the primers we added to the primer 1_DpnII a random sequence upstream the 8bp barcode and the MiniDs binding sequence-P1 frm [17]. We use identical barcodes for every generated library. Later in the sequencing procedure, they are differentiated by the attachment of a different barcode after the adapters⁶ are ligated to the DNA fragments of the library.

A details description of the Digestion, Ligation and PCR for the yLIC133 and yLIC136 samples is reported in <https://leilaicruz.github.io/Experimental-jupyterBOOK/journal/2020-08/2020-08-24-Digestion-circularization-PCR.html>

Sequencing. We specifically use Illumina sequencing [16, 49], which corresponds to the 2nd generation of next-generation sequencing technology. One of the advantages is that it is high through-

⁵From <https://sites.google.com/site/satayusers/complete-protocol/dna-sequencing>, used in [17]

⁶The adapters contain the sequencing primer binding sites, the index sequences, and the sites that allow library fragments to attach to the flow cell lawn. From <https://support.illumina.com/bulletins/2016/04/adapter-trimming-why-are-adapter-sequences-trimmed-from-only-the-ends-of-reads.html>



3

Figure 3.12. Steps towards DNA sequencing of a SATAY library. Inspired in protocol description from <https://sites.google.com/site/satayusers/complete-protocol/dna-sequencing>. Created with BioRender.com

put and suitable for whole-genome sequencing, yet it could be costly⁷. For this type of sequencing, transposon sequencing, you can apply two methods, either single end sequencing or pair end sequencing. The former involves sequencing the DNA from one end; it is the more straightforward form of Illumina sequencing, suitable for transposon sequencing as we are interested in detecting the insertion site downstream of the sequencing primer. We sequence both ends of a fragment and generate high-quality, alignable sequence data with the latter. Paired-end sequencing facilitates detection of genomic rearrangements and repetitive sequence elements, as well as gene fusions and novel transcripts⁸. We have used both types of methods, to sequence our generated libraries. The reasons are that the companies we chose did not give us an option to choose between one method or the other.

List of strains, plasmids and primers

⁷From <https://study.com/academy/answer/what-are-the-advantages-and-disadvantages-of-the-following-dna-sequencing-method-illumina-sequencing.html>

⁸From <https://emea.illumina.com/science/technology/next-generation-sequencing/plan-experiments/paired-end-vs-single-read.html>

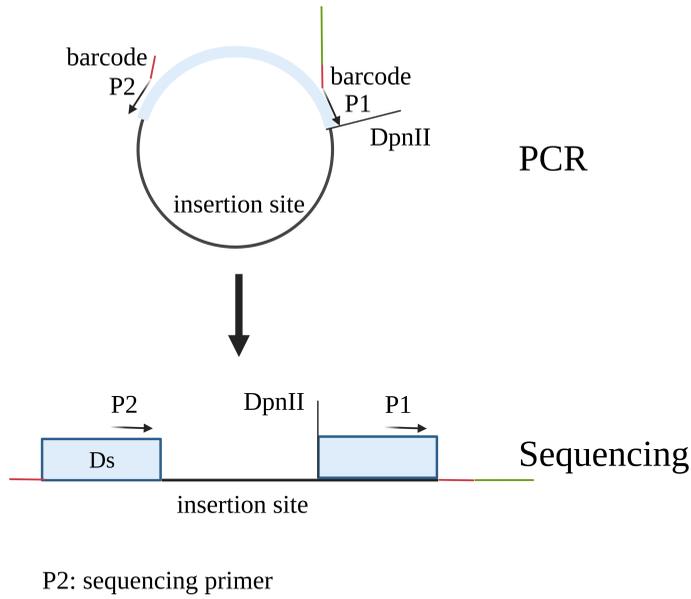


Figure 3.13. The insertion site is downstream of the sequencing primer. Hence the single end sequencing method in Illumina sequencing will be sufficient to get high-quality data.

Name	Genotype	Source
yII3a	MATx <i>can1-100,leu2-3,112</i> , <i>his3-11,15</i> , <i>ura3Δ</i> <i>BUD4-S288C</i>	Laan lab
yLIC133a	MATx <i>can1-100,leu2-3,112</i> , <i>his3-11,15</i> , <i>ura3Δ,ade2Δ</i> <i>BUD4-S288C</i>	This study
yLIC136a	MATx <i>can1-100,leu2-3,112</i> , <i>his3-11,15</i> , <i>ura3Δ,ade2Δ</i> <i>nrp1::HYGRO ,BUD4-S288C</i>	This study

Table 3.1. List of strains used in this chapter

Name	Source	Description
pBK549	[17]	MiniDs Transposons and Gal1-Transposase
pLL112	Laan lab	<i>URA3</i> fragment

Table 3.2. List of plasmids used in this chapter.

Name	Sequence	Source
olic15	GCTGTGtatggcactctcTTGATTG TTTTGTCCGATTT CTTGTTTTCTTG	This study
olic16	CAAGAAAAACAAGAAAAATCGG ACAAAACAATCAAG gaggtgcaccataCCACAGC	This study
olic17	GATGTAATCATAACAAA GCCTAAAA AATAGGTATATC GTGAGTTTAGTATA CATGCATTTACTTATAATACAG	This study
olic18	CTGTATTATAAGTAAATGCA TGTATACTAAA CTCACGATATACCTATT TTTTAGGCTTTGTTATGATT ACATC	This study
olic20	GATGTAATCATAACAAA GCCTAAAA AATAGGTATATCCTTG ATTGTTTTGTCGATTTTCTT GTTTTTCTTG	This study

Name	Sequence	Source
olic21	CAAGAAAAACAAGA AAATCGGAC AAAACAATCAAAGGAT ATACCTATTTTTAGGCT TTGTA TGATTACATC	This study
olic24	ATTACAGCTATGCTG ACAAATGACTCTTG	This study
olic26	GCTATCCTCGGTTCC TGCATTGAGC	This study
p1_DpnII	AGGTCAGTCACATGGTTAGG ACGCAGATAGACAGCCACATA acgaaaaacgaa ^{cgggataaa}	[17]
p2_DpnII	GCCACATATTTACCGACCGTTA CCGACCGTTTTTCATCCCTA	[17]
p3_NlaIII	AGGTCAGTCACATGGTTAGGACG CAGATAGACATAGGATGAcgaaaaacga acgggataaa	[17]
p4_NlaIII	TAGGATGATTTACCGACCGT TACCGACCGTTTTTCATCCCTA	[17]

Table 3.3. List of primers used in this chapter

Knockout of *URA3* and *ADE2* in *wild type*. We kick out the *ADE2* locus from the yLL3a strain following a two step procedure. First, the *URA3* genes was inserted in *ADE2* locus and then it was transformed by a empty sequence. The procedure consisted in making two DNA fragments (OEP1 and OEP2) that were transformed sequentially in yLL3a, using the lithium acetate high efficiency yeast transformation protocol [dx.doi.org/10.17504/protocols.io.gzrbx56](https://doi.org/10.17504/protocols.io.gzrbx56). The DNA fragments were produced by the method of overlap extension PCR, see [dx.doi.org/10.17504/protocols.io.psnndnd](https://doi.org/10.17504/protocols.io.psnndnd) for a whole description of this method . The OEP1 is the DNA fragment used to insert the *URA3* gene in the *ADE2* locus, and the OEP2 is the DNA fragment used to kick out the *URA3* gene from this location. For the OEP1, we need to produce three DNA fragments that are going to be glued together through PCR. The first DNA fragment corresponds to the upstream region of the *ADE2* gene, that was amplified by PCR. Specifically, for this PCR reaction, the template was genomic DNA from yLL3a and primers olic24/olic15 (I), the second to the *URA3* gene fragment amplified from the plasmid pLL112 with primers olic16 and olic17 (II). Lastly, the downstream region of *ADE2* which we obtained by PCR with the yLL3a genomic DNA and olic18/olic26 primers (III). Then, all three DNA fragments (I+II+III) were taken as template for the overlap extension PCR method with primers olic24/olic26. The OEP2 fragment consisted in gluing two DNA fragments with no homology to the *URA* gene. Firstly, the fragment from the yLL3a genomic DNA and primers olic24/olic20 and secondly the fragment from the same template and primers olic21/olic26. We checked the expected length of OEP1 and OEP2 by DNA electrophoresis and sequence by Sanger sequencing with Macrogen. This method produces the strain yLIC133.

Further, transformation of yLIC133 with an amplified DNA fragment from the yLL137 strain with primers oES83/oES84 yields the strain yLIC136.

Data analysis

The sequencing and data analysis of the *wild type* and *nrp1* full knockout mutant strain was performed by our collaborators and authors of the main publication on SATAY: Michel, A. H., Hatakeyama, R., Kimmig, P., Arter, M., Peter, M., Matos, J., ... and Kornmann, B. (2017). *Functional mapping of yeast genomes by saturated transposition*. *Elife*, 6, e23570. The sequencing and analysis of our strains was carried out in the facilities of Oxford University. The script for the analysis is published as a Supplemental file of the publication [50]: Michel, A. H., Kornmann, B. (2022). *Saturated Transposon Transposons Analysis in Yeast Yeasts (SATAY) for Deep Functional Mapping of Yeast Yeasts Genomes*. In *Yeast Functional Genomics: Methods and Protocols* (pp. 349-379). New York, NY: Springer US. In table 5.3 is shown the output of this analysis in terms of the number of transposons and reads annotated to the whole genome and to the coding sequences (CDS) of the different technical replicates of the *wild type* and *nrp1* mutants.

Preprocessing bash script to extract forward reads. This script is published in the following repository: https://github.com/SATAY-LL/comparative-analysis-among-strains/blob/main/publishing/scripts/forward_reads_extraction-new.sh

Transposonmapper. This workflow is created for processing sequencing data for Saturated Transposon Analysis in Yeast (SATAY) for *Saccharomyces Cerevisiae*. It is inspired by the Matlab code developed by [17]. Transposonmapper [51] has been developed in Bash and Python. It performs the steps from raw sequencing data until the transposon mapping that outputs files containing all insertion sites combined with the number of reads. The workflow requires input sequencing data in fastq format. It can perform the following tasks: sequence trimming, quality checking raw and trimmed fastq files sequence alignment with the reference genome (S288C

Strain	# of transposons	# of reads	# of transposons in CDS	# of reads in CDS
yLIC133_a	423524	23428274	194719	7246573
yLIC133_b	414347	24441674	189469	7577024
yLIC136_a	467104	28569602	287127	16494324
yLIC136_b	510327	25383712	301283	16057911

Table 3.4. Characteristics of the SATAY libraries of the wild type and *nrp1* mutant yeast strains, after the transposon sequencing data analysis. From left to right in the columns, the strain name, the number of transposons mapped, the number of reads mapped, the number of transposons mapped to coding sequences(CDS), and the number of reads mapped to CDS.

cerevisiae genome) quality checking bam files, indexing, and sorting transposon mapping. The output files indicate the location of transposon insertions and the number of reads at those locations. This is presented in both .bed and .wig format.

Transposonmapper is tailored to have a shareable package that other researchers can easily use and modify according to their needs. The package contains documentation for users and developers, a modular PyPI package for the postprocessing analysis (<https://pypi.org/project/transposonmapper/>). It has the option of installing the whole suite of Linux dependencies software if you are a developer, a test framework to maintain main functionalities, a docker container (<https://hub.docker.com/repository/docker/leilaicruz/satay/general>) to allow to run the software independent of the operative system. We also made a documentation website (<https://satay-ll.github.io/Transposonmapper/Introduction.html>) that describes all the features of the pipeline and how to use it. What is novel about this package is the possibility of accessing the code and customizing it to your needs due to its extensive documentation and test framework.

Fitness scripts. Fitness scripts can be found at the github repository: https://github.com/SATAY-LL/comparative-analysis-among-strains/blob/main/publishing/scripts/fitness_both_methods.py All the necessary files are located in the same folder. The scripts for the fig. 3.4 are also in the same script file.

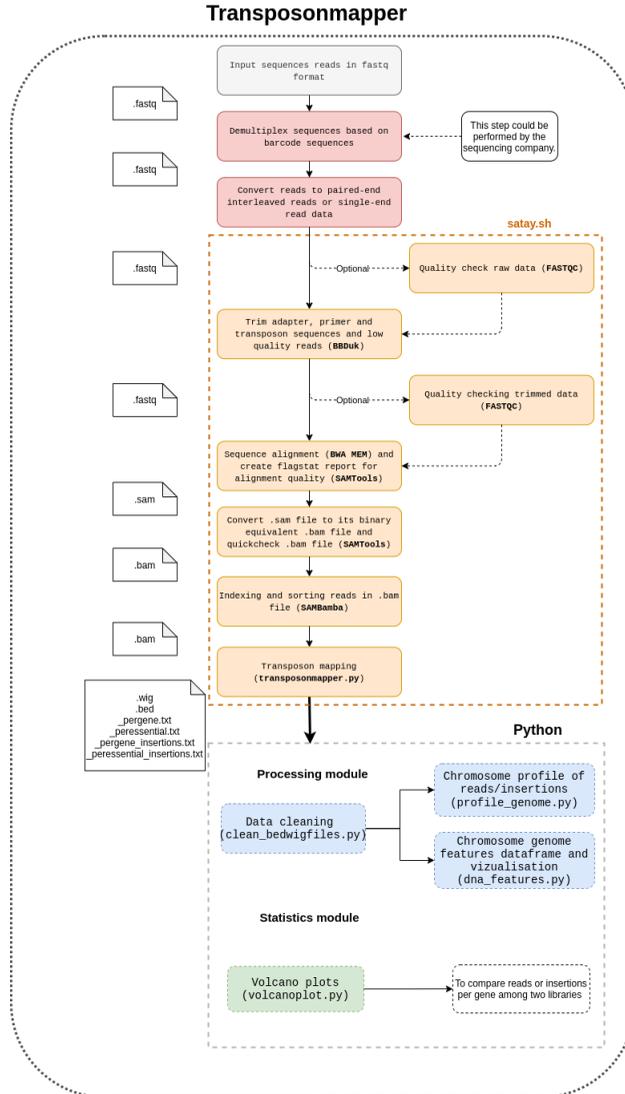


Figure 3.14. Flowchart of the main steps of the Transposonmapper pipeline.

Bibliography

1. Kingma, E., Dolsma, F., Inigo de la Cruz, L. M. & Laan, L. Saturated Transposon Analysis in Yeast as a One-step Method to Quantify the Fitness Effects of Gene Disruptions on a Genome-Wide Scale. *bioRxiv*, 2023-09 (2023).
2. Kuzmin, E. *et al.* τ -SGA: synthetic genetic array analysis for systematically screening and quantifying trigenic interactions in yeast. *Nature protocols* **16**, 1219–1250 (2021).
3. Liu, L. *et al.* Synthetic lethality-based identification of targets for anticancer drugs in the human signaling network. *Scientific Reports* **8**, 8440 (2018).
4. O'Neil, N. J., Bailey, M. L. & Hieter, P. Synthetic lethality and cancer. *Nature Reviews Genetics* **18**, 613–623 (2017).
5. Boone, C., Bussey, H. & Andrews, B. J. Exploring genetic interactions and networks with yeast. en. *Nature Reviews Genetics* **8**, 437–449. ISSN: 1471-0056, 1471-0064. <http://www.nature.com/articles/nrg2085> (2021) (June 2007).
6. Usaj, M. *et al.* TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. *G3 Genes | Genomes | Genetics* **7**, 1539–1549. ISSN: 2160-1836. <https://doi.org/10.1534/g3.117.040220> (2022) (May 2017).
7. Payen, C. *et al.* High-throughput identification of adaptive mutations in experimentally evolved yeast populations. *PLoS genetics* **12**, e1006339 (2016).
8. Hietpas, R., Roscoe, B., Jiang, L. & Bolon, D. N. Fitness analyses of all possible point mutations for regions of genes in yeast. *Nature protocols* **7**, 1382–1396 (2012).
9. Smith, A. M. *et al.* Quantitative phenotyping via deep barcode sequencing. *Genome research* **19**, 1836–1842 (2009).
10. Robinson, D. G., Chen, W., Storey, J. D. & Gresham, D. Design and analysis of Bar-seq experiments. *G3: Genes, Genomes, Genetics* **4**, 11–18 (2014).
11. Wetmore, K. *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly barcoded transposons, mBio 6 (2015). URL <https://mbio.asm.org/content/6/3/e00306-15>. Publisher: American Society for Microbiology Section: Research Article (2015).
12. Johnson, M. S., Martsul, A., Kryazhimskiy, S. & Desai, M. M. Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science* **366**, 490–493 (2019).
13. Venkataram, S. *et al.* Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast. *Cell* **166**, 1585–1596 (2016).
14. Cain, A. K. *et al.* A decade of advances in transposon-insertion sequencing. en. *Nature Reviews Genetics* **21**. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 9 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Bacterial genetics;High-throughput screening;Mobile elements;Mutagenesis Subject_term_id: bacterial-genetics;high-throughput-screening;mobile-elements;mutagenesis, 526–540. ISSN: 1471-0064. <https://www.nature.com/articles/s41576-020-0244-x> (2021) (Sept. 2020).

15. Van Opijnen, T., Bodi, K. L. & Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature methods* **6**, 767–772 (2009).
16. Van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends in genetics* **30**, 418–426 (2014).
17. Michel, A. H. *et al.* Functional mapping of yeast genomes by saturated transposition. en. *eLife* **6**, e23570. ISSN: 2050-084X. <https://elifesciences.org/articles/23570> (2021) (May 2017).
18. Rifat, D., Chen, L., Kreiswirth, B. N. & Nuermberger, E. L. Genome-wide essentiality analysis of *Mycobacterium abscessus* by saturated transposon mutagenesis and deep sequencing. *Mbio* **12**, 10–1128 (2021).
19. Michel, A. H., van Schie, S., Mosbach, A., Scalliet, G. & Kornmann, B. *Exploiting homologous recombination increases SATAY efficiency for loss- and gain-of-function screening* en. preprint (Cell Biology, Dec. 2019). <http://biorxiv.org/lookup/doi/10.1101/866483> (2021).
20. Lee, R. D. Population dynamics of humans and other animals. *Demography*, 443–465 (1987).
21. Bell, G. Experimental genomics of fitness in yeast. *Proceedings of the Royal Society B: Biological Sciences* **277**, 1459–1467 (2010).
22. Boone, C., Bussey, H. & Andrews, B. J. Exploring genetic interactions and networks with yeast. *Nature Reviews Genetics* **8**, 437–449 (2007).
23. Jacobs, M. A. *et al.* Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences* **100**, 14339–14344 (2003).
24. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *nature* **418**, 387–391 (2002).
25. Punta, M. *et al.* The Pfam protein families database. *Nucleic acids research* **40**, D290–D301 (2012).
26. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acids Research* **51**, D418–D427 (2023).
27. Wong, W. T. *et al.* A protein-binding domain, EH, identified in the receptor tyrosine kinase substrate Eps15 and conserved in evolution. *Proceedings of the National Academy of Sciences* **92**, 9530–9534 (1995).
28. Giaever, G. & Nislow, C. The yeast deletion collection: a decade of functional genomics. *Genetics* **197**, 451–465 (2014).
29. Mi, H. *et al.* The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic acids research* **33**, D284–D288 (2005).
30. Hunter, S. *et al.* InterPro: the integrative protein signature database. *Nucleic acids research* **37**, D211–D215 (2009).
31. Qian, W., Ma, D., Xiao, C., Wang, Z. & Zhang, J. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell reports* **2**, 1399–1410 (2012).
32. Kuzmin, E. *et al.* -SGA: synthetic genetic array analysis for systematically screening and quantifying trigenic interactions in yeast. en. *Nature Protocols* **16**, 1219–1250. ISSN: 1754-2189, 1750-2799. <http://www.nature.com/articles/s41596-020-00456-3> (2021) (Feb. 2021).
33. Cohen, I. *et al.* Pearson correlation coefficient. *Noise reduction in speech processing*, 1–4 (2009).

34. Van Leeuwen, J. *et al.* Exploring genetic suppression interactions on a global scale. *Science* **354**. Publisher: American Association for the Advancement of Science, aag0839. <http://www.science.org/doi/10.1126/science.aag0839> (2022) (Nov. 2016).
35. Teng, X. *et al.* Genome-wide consequences of deleting any single gene. *Molecular cell* **52**, 485–494 (2013).
36. Kim, D.-U. *et al.* Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nature biotechnology* **28**, 617–623 (2010).
37. Rousset, F. *et al.* The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nature microbiology* **6**, 301–312 (2021).
38. Chen, P., Michel, A. H. & Zhang, J. Transposon insertional mutagenesis of diverse yeast strains suggests coordinated gene essentiality polymorphisms. *Nature Communications* **13**, 1490 (2022).
39. Van Leeuwen, J., Boone, C. & Andrews, B. J. Mapping a diversity of genetic interactions in yeast. *Current opinion in systems biology* **6**, 14–21. ISSN: 2452-3100. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6269142/> (2022) (Dec. 2017).
40. Dowell, R. D. *et al.* Genotype to phenotype: a complex problem. *Science* **328**, 469–469 (2010).
41. Ralser, M. *et al.* The *Saccharomyces cerevisiae* W303-K6001 cross-platform genome sequence: insights into ancestry and physiology of a laboratory mutt. *Open biology* **2**, 120093 (2012).
42. Wong, E. D., Karra, K., Hitz, B. C., Hong, E. L. & Cherry, J. M. The YeastGenome app: the *Saccharomyces* Genome Database at your fingertips. *Database* **2013** (2013).
43. Bosch-Guiteras, N. & van Leeuwen, J. Exploring conditional gene essentiality through systems genetics approaches in yeast. *Current Opinion in Genetics & Development* **76**, 101963 (2022).
44. Lazarow, K., Du, M.-L., Weimer, R. & Kunze, R. A Hyperactive Transposase of the Maize Transposable Element *Activator (Ac)*. en. *Genetics* **191**, 747–756. ISSN: 1943-2631. <https://academic.oup.com/genetics/article/191/3/747/5934995> (2021) (July 2012).
45. Weil, C. F. & Kunze, R. Transposition of maize *Ac/Ds* transposable elements in the yeast *Saccharomyces cerevisiae*. *Nature genetics* **26**, 187–190 (2000).
46. Wikipedia contributors. *Homologous recombination* — *Wikipedia, The Free Encyclopedia* [Online; accessed 14-February-2022]. 2022. https://en.wikipedia.org/w/index.php?title=Homologous_recombination&oldid=1069919973.
47. Wiktionary. *pericentromeric* — *Wiktionary, The Free Dictionary* [Online; accessed 15-February-2022]. 2018. <https://en.wiktionary.org/w/index.php?title=pericentromeric&oldid=49476016>.
48. Levitan, A. *et al.* Comparing the utility of in vivo transposon mutagenesis approaches in yeast species to infer gene essentiality. en. *Current Genetics* **66**, 1117–1134. ISSN: 0172-8083, 1432-0983. <http://link.springer.com/10.1007/s00294-020-01096-6> (2021) (Dec. 2020).
49. Ambardar, S., Gupta, R., Trakroo, D., Lal, R. & Vakhlu, J. High throughput sequencing: an overview of sequencing chemistry. *Indian journal of microbiology* **56**, 394–404 (2016).
50. Michel, A. H. & Kornmann, B. in *Yeast Functional Genomics: Methods and Protocols* 349–379 (Springer, 2022).

51. Iñigo de la Cruz, L., van Beek, G. & Kok, M. *transposonmapper* version 1.1.5. <https://github.com/SATAY-LL/Transposonmapper>.

Using genetic interactions from saturated transposition analysis to uncover *NRP1*'s function in *S.cerevisiae*.

No great discovery was ever made without a bold guess.

Sir Isaac Newton

Abstract Cellular polarity stands as a fundamental process governing both cell division and morphogenesis. The establishment of polarity marks the initiation of cell division. Despite extensive studies dissecting the functionality of cellular polarity network components, numerous proteins associated with polarity still harbor unknown functions. This chapter explores the biological role of a polarity-associated protein in the yeast *S. cerevisiae*, named Nrp1. While Nrp1 has been observed to restore polarity establishment in $\Delta bem1$ mutants, the precise biological mechanisms underpinning this recovery remain unclear. This study adopts a two-fold approach to investigate the function of *NRP1*, which encompasses 1) quantifying whether the RNA binding domain annotated to Nrp1 influences gene expression in different mutants, and 2) identifying genome-wide genetic interaction partners through SATAY, aiming to propose potential biological functions involving *NRP1*. We did not find a significant influence on gene expression upon *NRP1* full knockout from the wild-type and $\Delta bem1$ backgrounds. The enrichment of predicted genetic interactors from SATAY suggests that *NRP1* is linked in processes related to prion formation, endocytosis, and cell cycle checkpoints. The same analysis with annotated polarity genes hints that *NRP1* deletion rescues *bem1* mutants by either regulating the precise cell size at the START point of the cell cycle or enhancing the Cdc42 exchange from the membrane to the cytosol. Further, specific low-throughput experiments are proposed to test the mechanistic hypotheses regarding the function of Nrp1 in yeast cells.

4.1. Introduction

Cellular polarity is a fundamental process that governs cell division and morphogenesis, playing a crucial role in the development of various organisms. It involves critical cellular processes such as cell division, differentiation, migration, signaling, and fertilization.

The yeast species *Saccharomyces cerevisiae* serves as an exemplary model to explore the molecular mechanisms underlying cell polarity. This organism demonstrates prominent polarization in response to both intracellular and extracellular cues. Moreover, yeast cells share common features with other organisms concerning cell polarity, including the regulation by intrinsic and extrinsic cues, the presence of conserved key regulatory molecules like small GTPases, and the occurrence of asymmetric growth of the cytoskeleton [1, 2].

In *Saccharomyces cerevisiae*, cell polarity is evident during vegetative growth through budding, which is guided by landmark proteins as intrinsic cues. Additionally, cell polarity is observed during mating between cells of opposite mating types and during filamentous growth induced by nutrient deprivation, such as nitrogen limitation. However, our study focuses on the first type of cell polarity, which involves the asexual reproduction of yeast cells through budding. We aim to investigate the process of polarity establishment as the initial and critical step in this reproductive process.

The polarity establishment system in budding yeast has been extensively studied, with abundant literature available on the various proteins involved and their roles within the system [1–4]. Protein-protein interaction studies and genetic analyses have allowed the representation of the system as a network, where proteins are nodes and their interactions are depicted as edges [5].

The establishment of cell polarity occurs at the end of the G1 phase of the cell cycle, panel A) fig. 4.1, and it defines the beginning of the cell cycle. At this point, certain checkpoints related mainly to nutrient availability and protein levels must be successfully cleared to progress in the cell cycle. Subsequent phases such as S (Synthesis), G2 (Growth), and M (Mitosis) are primarily responsible for DNA synthesis, proper DNA duplication, and chromosome segregation, respectively, in the new daughter cell.

We conceptualize the polarity establishment function as an interconnected module that interacts with mainly five different pathways [5] that occur hierarchically in the cell. These pathways are the mating, bud scar, reaction-diffusion, actin, and timing pathways, as depicted in fig. 4.1. The timing pathway is mainly involved in the cell cycle cues that shape the temporal regulation of all processes in the cell.

An evolutionary study [6] investigated the adaptation of the polarity establishment network after removing the near essential scaffold protein Bem1. Cells lacking Bem1 experienced a considerable fitness impairment, struggling to form buds and displaying an extremely low survival rate. However, these cells restored their fitness levels through a 1000-generation evolutionary serial dilution experiment.

Surprisingly, the evolved cells achieved polarization similar to the wild-type phenotype but through a novel alternative mechanism that involved the absence of three additional proteins: Bem3, Nrp1 (panel B, fig. 4.1), and Bem2. Bem3 and Bem2 are recognized Guanine Activating Proteins (GAPs) for the GTPase Cdc42. They facilitate the hydrolysis of GTP (Guanosine triphosphate) bound in its active state into GDP (Guanosine diphosphate), transitioning it to an inactive state. The study by Brauns et al. [7] investigated the detailed mechanisms through which *BEM3* functions as a significant repressor in $\Delta bem1$ mutants by significantly restoring the ability to polarize of such mutants. *NRP1* knockout in *bem1bem3* mutants was observed, in the same evolutionary

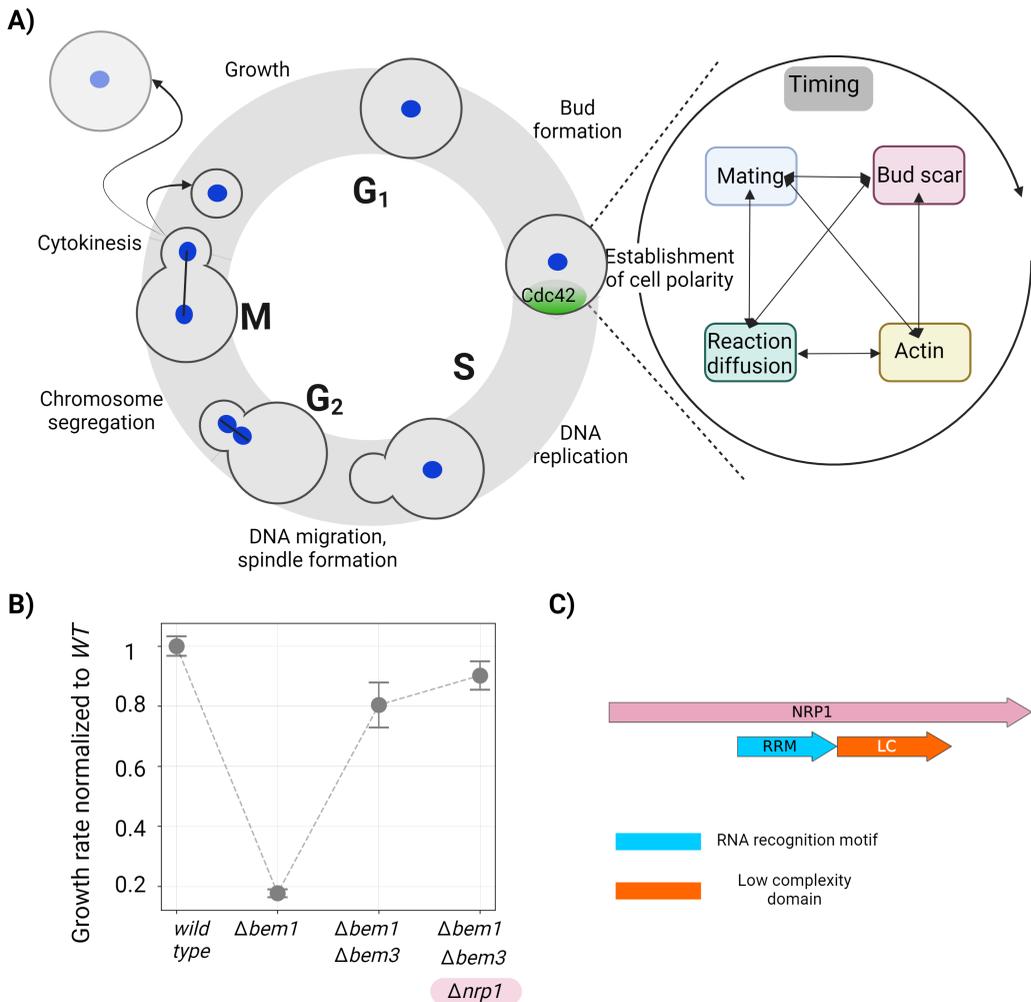


Figure 4.1. Polarity establishment and Nrp1 state of the art. **A)** Polarity establishment in budding yeast cells as part of the cell cycle and as the interplay of multiple pathways. The interdependence between the mating, the bud scar, the reaction-diffusion, the actin, and the timing pathways results in accumulating the GTPase Cdc42 bound to GTP molecules (i.e., active Cdc42) to one point on the plasma membrane, which can signal downstream effectors to proceed the cell cycle, at the end of the G₁ phase of the cell cycle. **B)** Growth rate recovery of $\Delta bem1$ gene knockout by *BEM3* and *NRP1* knockouts. **C)** Current knowledge on *NRP1* sequence indicates that the Nrp1 protein contains an RNA recognition motif and low complexity domains.

experiment, to increased their fitness, panel B fig. 4.1.

Notably, Nrp1 has not been previously associated with cell polarity in any other study. Hence, the specific mechanisms through which the deletion of *NRP1* rescues impaired $\Delta bem1$ mutants remain unknown, making it an intriguing aspect that motivates further investigation of the function of Nrp1 in budding yeast.

Importantly, Nrp1 exhibits high conservation across the fungal kingdom, particularly within the Ascomycota phylum [8]. This level of conservation is comparable to the prevalence of Bem1

within the same phylum, suggesting that Nrp1 likely possesses a function of significant importance in yeast evolution. The conservation of Nrp1 across fungal species further emphasizes the potential significance of studying its role and contribution to cellular processes in yeast.

The gene sequence of *NRP1* [9, 10] consists of two distinct domains: an RNA recognition motif (RRM) and a low complexity domain containing repeated regions rich in asparagine residues (panel C) fig. 4.1). The presence of the RRM suggests that Nrp1 has the potential to bind and regulate mRNA molecules. However, the specific mRNA sequences that interact with Nrp1 have not been identified yet [11]. Moreover, it has been suggested to be involved in ribosome biogenesis [9, 12, 13].

4.2. Results

4

4.2.1. The RNA binding domain of *NRP1* does not strongly regulate gene expression in wild-type and Δ *bem1* mutants.

To investigate the potential influence of Nrp1 knockout on protein translation and gene expression through its RNA recognition motif, we conducted mass spectrometry analysis [14] on various mutants of yeast cells.

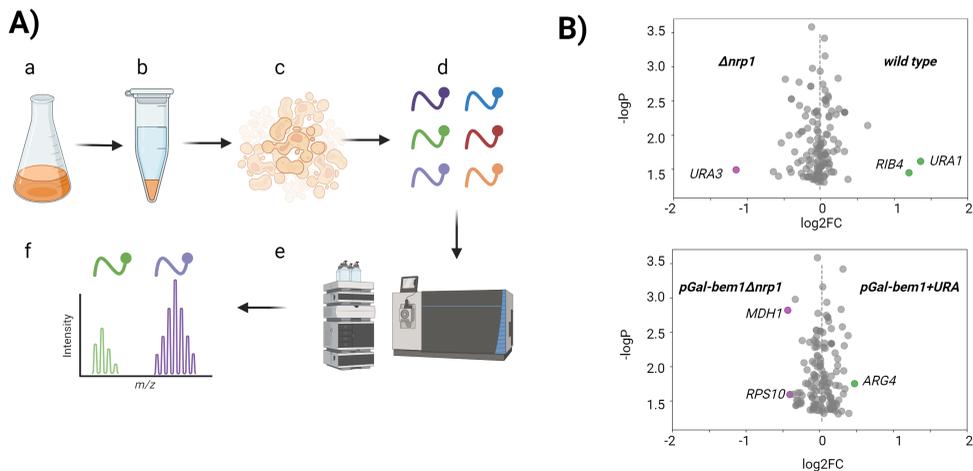


Figure 4.2. Mild changes in gene expression upon Nrp1 deletion in different mutants. **A)** A typical workflow for a shotgun mass spectrometry involves yeast cell lysis, protein extraction, and proteolytic digestion (a-c); proteins are enzymatically hydrolyzed into peptides and purified (d) followed by liquid chromatography (LC) coupled to a hybrid mass spectrometer capable of isolating intact ions, to measure their masses (e). Next, thousands of spectra are collected for qualitative analysis of peptides in a sample. The first data analysis step identifies the peptides and proteins from the complex sample mixture (f). **B)** Volcano plots representing differences in protein abundances across different mutants. (top) Δ *nrp1* vs *wild type* (down) *pGal-bem1* Δ *nrp1* vs *pGal-bem1*. Genes to the left of the plot ($\log FC < 0$) are genes that are upregulated in the mutant background relative to the *wild type* expression level. Likewise, genes with positive values of $\log FC$ correspond with genes downregulated in the mutant compared to the *wild type*. The top dashed arrow indicates the mean of the distribution of gene expression of all protein values, indicating a measure of which effect is more predominant concerning the values of the gene expression fold change.

For this analysis, we employed a shotgun mass spectrometry approach, a commonly used method for unbiased protein quantification in biological samples [15]. The detailed protocol

followed for the experiment can be found in the Materials and Methods section, specifically in section section 4.5.

To assess the impact of Nrp1 on gene expression through its RNA binding motif, we analyzed the differences in protein abundances upon Nrp1 knockout in both the *wild type* background and the $\Delta bem1$ mutants. The data analysis was performed using the PEAKS software [16], which detected 1277 proteins across all mutants, accounting for approximately 20% of the yeast genome size. Among these proteins, only 147 showed significant changes in abundance across the mutants, as determined by statistical analysis (p -values < 0.05) carried out with the software PEAKS.

In our investigation, we compared protein abundances across different genetic backgrounds to gain insights into the role of Nrp1 in gene expression. We calculated the log fold change (logFC) for each protein's abundance between two backgrounds of interest. Panel **B**) of fig. 4.2 displays the results in a volcano plot.

A notable observation is that the *URA3* gene, which is constitutively expressed in the $\Delta nrp1$ strain, shows a higher abundance compared to the *wild type* strain, which has an *ura3* auxotrophy. This finding highlights the significance of auxotrophic markers like *URA3*, which can exert considerable effects on the cell under specific conditions, as observed in the $\Delta nrp1$ mutants.

Additionally, we noticed that the genes *RIB4*¹ and *URA1* are overregulated in the *wild type* background. This observation suggests the possibility that the expression of *URA3* actively inhibits the expression of *URA1*. Ura1 and Ura3 enzymes are involved in the biosynthesis of pyrimidines in the cell.

With the observation that Nrp1 knockouts do not lead to significant changes in gene expression in both *wild type* and *bem1* mutant backgrounds, we turn our focus to exploring the global interaction network of Nrp1. For this purpose, we will employ Saturated Transposition Analysis in Yeast (SATAY).

4.2.2. The fitness effects distribution in the $\Delta nrp1$ genetic background is skewed.

Disabling the *NRP1* gene does not appear to have a significant effect on the fitness of wild-type yeast cells, as observed through measurements of growth rate and morphology, in our lab. Thus, studying its biological function from the changes it causes to phenotype in wild-type cells is unfeasible. Therefore, we aim to find which other mutations may perturb the phenotype (i.e., growth rate values) of $\Delta nrp1$ yeast mutants, which will shed light on which molecular pathways or biological functions are influenced by the removal of *NRP1*.

To explore the changes to the fitness that multiple single gene knockouts can cause to $\Delta nrp1$ yeast mutants, we utilize the capabilities of SATAY to generate a genome-wide library of double gene deletions ($\Delta nrp1\Delta genex$) using transposons (panel A) fig. 4.3) that covers nearly all yeast coding sequences. We use the method described in chapter 3 to generate relative fitness values. Considering the scarce knowledge regarding Nrp1's biophysical role, the distribution of fitness effects of $\Delta nrp1$ yeast mutants is taken as a first step to investigate which biological pathways are *NRP1* related.

In this regard, we confirmed that the predicted fitness values obtained from SATAY give consistent results compared to previous in-house measurements using population growth experimental procedures, see fig. 4.11 in different yeast mutants. The distribution of fitness effects (DFE) of gene knockouts in the $\Delta nrp1$ background, panel B) fig. 4.3, is normalized such as it is centered on the relative fitness value of $\Delta nrp1$ mutants from the SATAY wild-type library. The

¹Rib4 is an enzyme that is involved in riboflavin biosynthesis [17]

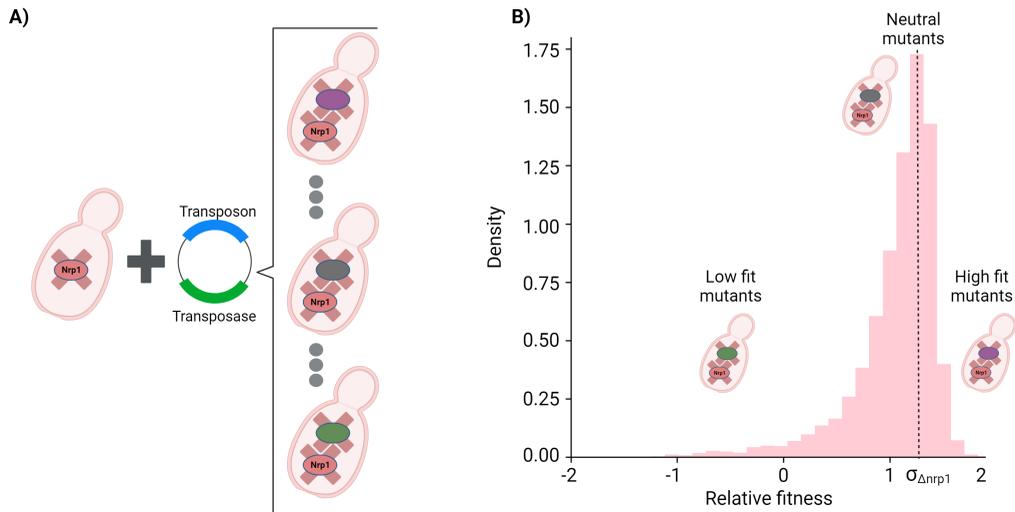


Figure 4.3. The distribution of fitness effects on $\Delta nrp1$ yeast mutants from SATAY is left skewed. **A)** SATAY on $\Delta nrp1$ yeast mutant generates a multiple yeast double knockout library. **B)** Distribution fitness effects of the generated double yeast knockout library, using SATAY. The dashed line represents the median of the distribution, which is the relative fitness of $\Delta nrp1$ mutants from the wild-type single knockout SATAY library.

predicted fitness, in terms of growth rate, from SATAY, is around the relative value of 1.2 compared to the median value of the wild-type library. The DFE is left-skewed. Thus, there are more fitness differences toward low fitness values.

We subsequently explore whether the genes that show a change of gene expression after *NRP1* knockout are correlated with their fitness values in the $\Delta nrp1$ background.

4.2.3. The fitness effects of gene knockouts are uncorrelated with the change of gene expression of the same genes in $\Delta nrp1$ mutants

The coupling between gene expression and fitness can provide insights of gene expression changes as an adaptive mechanism. We hypothesize that genes that are down-regulated in a specific background are prone to be suppressors in that background, in contrast to genes that are up-regulated, for which we predict its knockout could have detrimental effects.

We study this correlation with the changes on gene expression in $\Delta nrp1$ mutants compared to wild type, and their fitness values, inferred from SATAY, in the $\Delta nrp1$ genetic background. We observed that the genes up and down-regulated in wild-type background by the *NRP1* deletion exhibit similar mean fitness values close to 1, which indicates that those genes show neutral fitness effects in the wild-type background, fig. 4.4.

In summary, our analysis does not reveal clustering of down-regulated genes with high fitness values, nor upregulated genes with low fitness values. One possibility is that the observed changes in gene expression are not substantial enough to account for extreme alterations in fitness associated with the knockout of those genes.

In the next section, we explore the main biological functions associated with *NRP1*, which are derived from a functional enrichment of the predicted significant genetic interactors of *NRP1* from the method shown in chapter 3 to derive genetic interactions from SATAY read-outs.

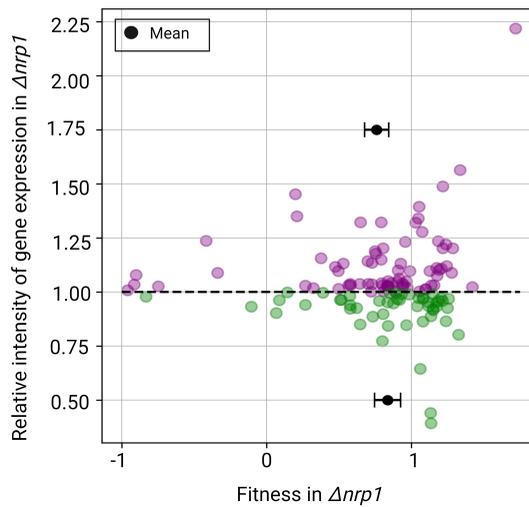


Figure 4.4. No correlation between up and down-regulated genes with their knockout fitness in $\Delta nrp1$ background. The y-axis represents the ratio of the relative protein abundances to the *wild-type* level, and the x-axis displays the predicted SATAY fitness of a mutant carrying a loss-of-function mutation in the protein of interest in the $\Delta nrp1$ background. The black dot on both sides of the plot symbolizes the mean of the fitness for up and down-regulated genes, and the error bars are the standard deviation.

4.2.4. Prion formation, endocytosis and cell cycle checkpoints related processes stand up as the most enriched functions for SATAY predicted *NRP1* genetic interactors.

In this section, our main objective is to investigate the genetic interactors of Nrp1 at a genome-wide scale to dig into the biological pathways associated with Nrp1. We will use the fitness values from the SATAY libraries in both the *wild type* and *nrp1* genetic backgrounds to compute the genetic interactors. The process of calculating genetic interactions from fitness values is described in section 3.2.4.

In short, to identify genetic interactions, we apply a multiplicative model to compare the double knockout relative fitness values with the expected relative fitness values, assuming that the fitness effects of single gene knockouts are independent and multiplicative. If the double knockout fitness is significantly higher than expected, it indicates a positive interaction, suggesting that both genes may suppress or mask each other's effects. Conversely, if the double knockout fitness is significantly lower than expected, it indicates a negative interaction between the pair of genes. The extreme case of a negative interaction is called synthetic lethality, where the double knockout is lethal, but every single knockout is viable, see fig. 4.5A.

In particular, the genetic interaction scores are computed following eq. (4.1):

$$\epsilon_{i-nrp1} = f_i|\Delta nrp1 - f_i|WT \times f_{nrp1}|WT \quad (4.1)$$

The fitness computation is performed using the procedure for the average fitness along each gene, explained in chapter 3. In fig. 4.11, panel C), we perform a normalization of $f_i|\Delta nrp1$ such that its median matches $f_{nrp1}|WT$. This normalization assumes that the probability of interaction

between a random gene i and our gene of interest, in this case, *NRP1*, is zero. Consequently, we ensure that our genetic interaction score distribution for all genes is centered around zero.

Furthermore, we apply a T-test [18] for the means of two independent sample scores. This test quantifies the difference between the arithmetic means of the two samples and provides p-values that indicate the significance of the difference between the samples. A p-value > 0.5 suggests that the observed difference in fitness values is likely due to chance, and, therefore, we discard those observations. The choice for significant p-values in these datasets was set at a threshold below 0.1. Panel D of fig. 3.7 illustrates the variance in fitness values among biological replicates of the strain carrying a knockout mutation in *NRP1*.

All computed genetic interaction scores for *NRP1* are represented through a volcano plot shown in fig. 4.5. The x-axis represents the interaction score, and the y-axis displays the statistical significance of each interaction as $-\log p$ -value. The plot resembles the shape of a volcano, hence its name. In this plot, we can observe extreme changes in both positive and negative interactors that are statistically significant, as determined by their t-statistic and fold change values. Positive values correspond to genes whose knockout suppresses the *nrp1* phenotype, while negative values represent genes whose removal exacerbates the *nrp1* phenotype.

To ensure that the differences in fitness between the double mutant and the product of the single mutants are significantly different, we applied a stringent cutoff to classify genetic interactions, rather than simply considering genes with scores different from zero, as shown in eq. (4.1). We employed a cutoff based on one standard deviation from the mean score value, shown in eq. (4.2). This approach allows us to identify genetic interactions that are more likely to be biologically meaningful and not just due to random variation.

$$\epsilon^+ > \mu_\epsilon + \sigma_\epsilon; \quad \epsilon^- < -\mu_\epsilon - \sigma_\epsilon \quad (4.2)$$

In the volcano plot of fig. 4.5, purple dots represent statistically significant positive genetic interactions, while green dots represent statistically significant negative ones. Our analysis identified 47 genes that show significant positive interactions with *NRP1* and 44 other genes that exhibit significant negative interactions with *NRP1*. These genetic interactors are likely to be functionally linked to *NRP1* and may play important roles in the biological pathways involving this gene.

To identify which biological pathways *NRP1* may be related to, we conduct a functional enrichment analysis of the genetic interactors of *NRP1*. Previous studies have suggested that suppressing interactions may involve proteins that function in a linear pathway or are part of a non-essential protein complex [20, 21]. On the other hand, negative genetic interactions may indicate redundant roles, suggesting that the studied gene pair is involved in an essential cellular function through complementary pathways.

The results of our analysis reveal that significant positive genetic interactions are enriched for a wide variety of biological processes, namely: inositol phosphate-related metabolic processes, cellular response to ethanol, mitotic spindle checkpoint, regulation of GTPase activity, and dephosphorylation, see panel C) fig. 4.5. Furthermore, the phenotypes related to the identified positive interactions are related to the respiratory growth phase of the yeast population growth, the exit from GO (stationary phase), spindle morphology, budding pattern, and RNA localization, among others, see panel D) fig. 4.5.

On the other hand, significant negative interactors are enriched for biological processes such as hydrogen peroxide and nucleotide sugar metabolic processes, spindle pole body duplication,

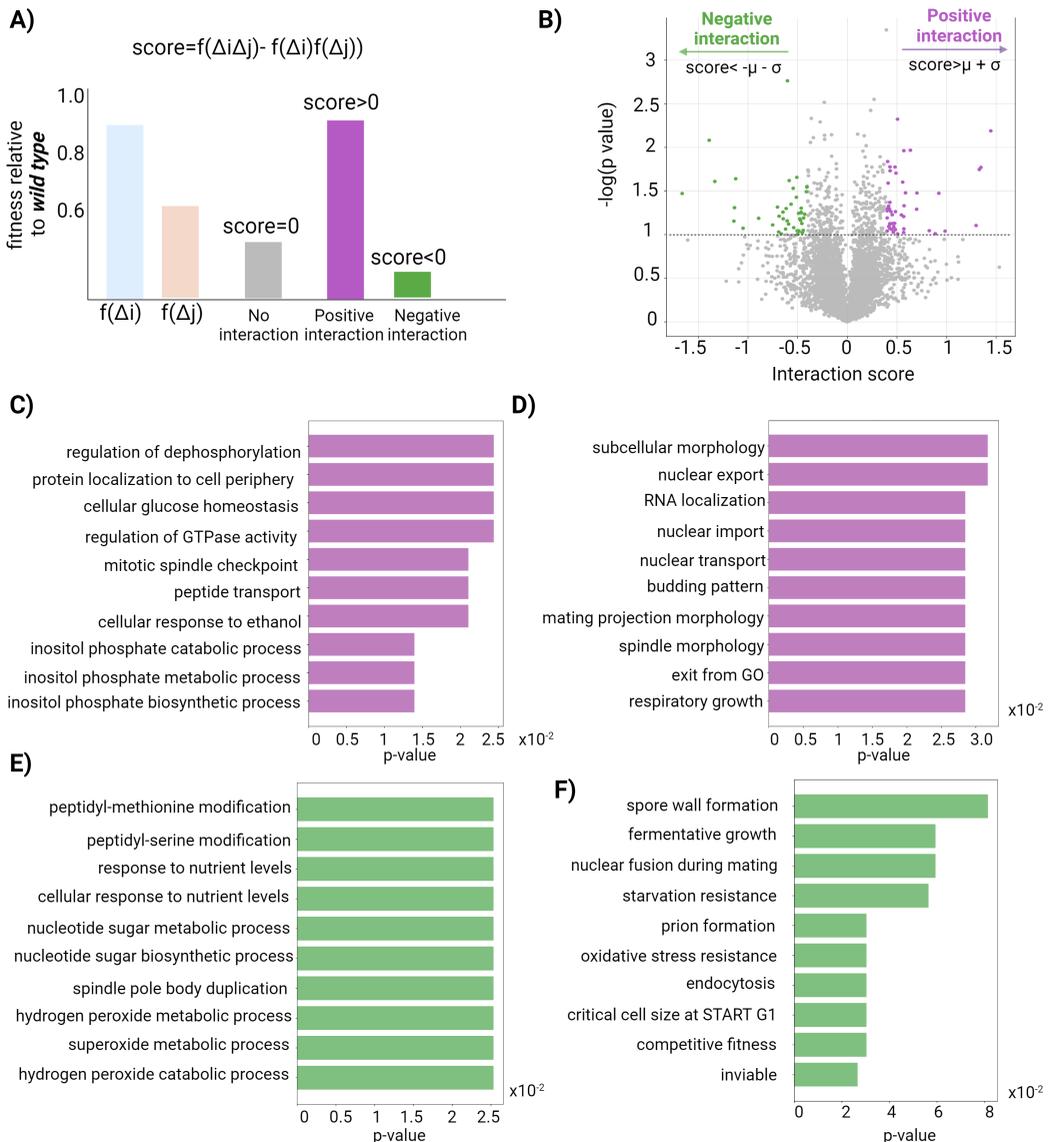


Figure 4.5. Identification and functional enrichment analysis for the genetic interactors of *NRP1*. **A)** Genetic interactors are defined as gene pairs whose combined perturbation gives rise to an unexpected phenotype. Suppose the fitness of the double knockout largely increases from the product of the single mutant fitness. In that case, it will indicate a positive interaction, while a noticeable reduction will hint at a negative interaction. **B)** Volcano plot of the genetic interactors for *NRP1*. The x-axis displays the interaction score computed by subtracting the product of the single knockout fitness from the double knockout fitness. This score is the average among the two computed scores from each library from *wild type* and $\Delta nrp1$. Purple and green dots represent significant positive and negative genetic interactions, respectively. **C)** Functional enrichment of significant positive genetic interactors for biological processes using the python package Gseapy [19]. **D)** Functional enrichment in phenotypes associated with knockout mutations of the significant positive interactors. **E)-F)** Same as **C)-D)** but on the significant negative genetic interactors.

cellular response to nutrient levels, and peptidyl-methionine and serine modifications, see panel E) fig. 4.5. Moreover, significant negative interactors are also enriched for phenotypes related to critical cell size at START in G1, endocytosis, oxidative and starvation stress resistance, prion formation, nuclear fusion during mating, fermentation growth, and spore wall formation, panel F) fig. 4.5.

Taken all together, *NRP1* seems to be linked in a complex interaction network involving metabolic stress, and cell cycle checkpoints-related genes. Specifically, looking at the common enrichment among both types of interactors, *NRP1* seems to be involved in the regulation of the first part of the cell cycle comprising the G0 and G1 phases and part of the G2 phase at regulating spindle pole body duplication and morphology. Furthermore, this analysis renders *NRP1* as a candidate that regulates the fermentation and respiratory phases from the population yeast growth model [22]. This is also backed up by the links from both types of interactors to processes related to the cellular response to ethanol and nutrient levels. Remarkably, this has been previously observed experimentally in mutants involving a double knockout mutation on the genes *BEM3* and *NRP1* by [23].

It is worth to mention that this study should be taken as the first step towards uncovering the biochemical pathways that *NRP1* may be impinging on because we lack two additional important measures: 1) correlation interaction profiles among the targeted genes to add information that those genes (especially the positive interactors) are part of the same physical complex or pathway [24]. 2) Low throughput experimental checks on the predictions of this approach.

In the next section, we ask to what extent these results from SATAY cover the existing knowledge about the genetic interactors of *NRP1* from other high throughput assays.

4.2.5. Predicted significant genetic interactors do not match existing *NRP1* genetic interactors

In addition to identify novel biophysical functions that *NRP1* may be linked to, we would like to study whether the predicted functions enclose or are a subset of known functional relationships that involve *NRP1*. We compare our predicted significant interactors with existing genetic interactors of *NRP1* from the Saccharomyces Genome Database (SGD) [17]. Specifically, we look at gene deletions that suppress or exacerbate the *NRP1* gene deletion phenotype².

We found 16 previously identified *NRP1* genetic interactions from [26] study, 8 of them positive and the rest negative interactors. We did not find any match from the 91 significant genetic interactors that we found using SATAY. However, for the eight existing positive interactors, we recovered one of them with a moderate p-value of 0.21 and a genetic interaction score of 0.25. The gene is *SYF1*, which its gene product is a mRNA-splicing protein that is part of the NineTeen Complex (NTC) and U2-type catalytic spliceosome; localizes to both nucleus and cytosol[17]. *SYF1* is currently annotated as an essential gene in *wild type* cells [27].

Out of the eight previously found negative interactors, we recovered two of them, with a statistical significance of 0.15 and genetic interaction score of -0.36. Those genes are *REE1*³ and *POP4*⁴.

²To do this, we impose a directional interaction, we considered the interactors classified as "Hit" in the database, where the 'rescuer' gene is annotated as the hit and the 'rescued gene' as the bait [25]. Thus, we focused on genes reported as rescuers of *NRP1* phenotypes.

³A protein whose biological role is unknown; localizes to the cytoplasm in a large-scale study [17]

⁴Subunit of both RNase MRP and nuclear RNase P; RNase MRP cleaves pre-rRNA, while nuclear RNase P cleaves tRNA precursors to generate mature 5' ends and facilitates turnover of nuclear RNAs; binds to the RPR1 RNA subunit in RNase P

The discrepancies between the predicted genetic interactors from SATAY and the existing ones are attributed to the differences in fitness approaches used in these methods, as illustrated in chapter 3, fig. 3.9. On the other hand, the genetic backgrounds of the mutated yeast strains and growth environments are distinct among these techniques. In addition, the noise sources during the SATAY data production, regarding the PCR and sequencing steps, add a considerable variation among the replicates of our experiment, which directly influences the statistical significance of the genetic interaction score values.

Despite these effects, we will rely on the genetic interactors predicted with SATAY to uncover relevant functions associated with *NRP1* since we have confirmed the repression effect of deleting *NRP1* in *bem1* mutants, panels B and D fig. 4.11, the neutrality of *nrp1* phenotype, panels B and F fig. 4.11 and the fitness defect in *bem1* mutants through SATAY, panels C and D fig. 4.11. However, it is essential to acknowledge that all the results obtained in this study are working hypotheses that warrant further experimental validation and testing.

4.2.6. Eleven annotated essential genes may have lost their essentiality after a *NRP1* knockout.

Extreme genetic interaction cases tell us about the possible genetic rewiring or reorganization when, in this case, *NRP1* is removed from the network. For that, we look at the genes that possess the strongest interactions with *NRP1* according to our experimental assay and data analysis procedure. Specifically, we ask if there are cases of annotated essential genes that lost the essentiality to become significant positive interactors of *NRP1* and vice-versa genes that seem to become essential or synthetic lethal in $\Delta nrp1$ genetic background.

We identified eleven annotated essential genes that seem to have lost their essential role in *nrp1* knockouts. All of these genes exhibit positive interaction scores above the threshold (eq. (4.2)) and p-values below 0.1. To further validate whether these genes may indeed be essential under our wild-type conditions, we assessed whether their fitness values fall below the median fitness value of essential genes in the wild-type background (fig. 4.6A). This step ensures that these genes become dispensable upon *NRP1* knockout in our experimental conditions. We applied the same fitness threshold for the mutant background, assuming that the likelihood of annotated essential gene knockouts increasing or decreasing their fitness in the mutant background is equally probable.

Thus, we selected genes that not only exhibit significant positive interactions with *NRP1* but also have a fitness value greater than 0.56 in the mutant background and below this value in the wild-type background. Only six genes out of the eleven met these criteria, although two were on the threshold, which we consequently excluded (see upper left square in panel B of fig. 4.6). These genes are: *INO2*, *MEC1*, *TFB4* and *ARB1*.

INO2 is a transcription factor that activates the transcription of genes involved in phospholipid metabolism in response to inositol depletion[28], and it is also involved in the diauxic shift[29]. Furthermore, it has been shown to affect nuclear segregation and bud pattern formation [30]. The insertion map of this gene in wild-type background shows accordingly a large region void of transposons, mainly downstream the gene (panel C) fig. 4.6). In the mutant background, the transposon insertions and read counts over the gene extend to a larger genomic area. This may indicate that, from the functional viewpoint, the primary function of *INO2* is being buffered in $\Delta nrp1$ mutants, which could be due to: 1) The inositol synthesis in $\Delta nrp1$ mutants is never

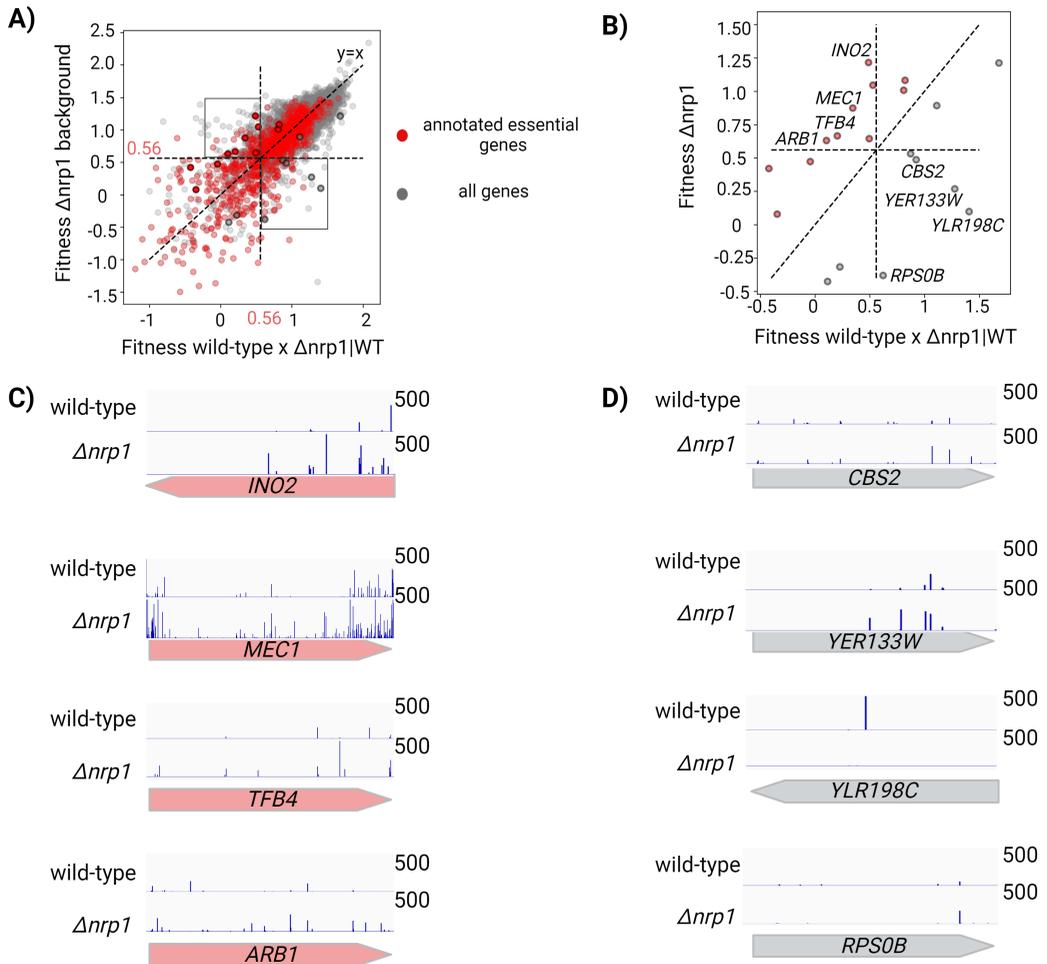


Figure 4.6. Predicted genes that lost and acquired an essential role in $\Delta nrp1$ mutants. **A)** Fitness values in $\Delta nrp1$ background against fitness values in wild type background. The median of the fitness values for annotated essential genes in our wild-type background is 0.56. Red and gray dots indicate annotated essential and all genes, respectively. We choose the same cut-off for the $\Delta nrp1$ background. We select genes whose fitness in $\Delta nrp1$ background is greater than 0.56 and below this value in the other background. In addition, we only select the genes that have a p -value < 0.1 concerning their interaction score with *NRP1*. The upper left square encloses the annotated essential genes that lost their essentiality, and the right bottom square encloses the predicted new essential genes in the mutant background. **B)** A zoom-in on the genes lying inside the squares of the figure from panel A). The selected genes based on their fitness values and significance have their name on the data point. **C)** Transposon insertion maps of the predicted genes that have lost essentiality in $\Delta nrp1$ mutants in both genetic backgrounds. The height of the bar indicates the number of reads associated with that insertion location. **D)** Transposon insertion maps of the predicted genes that acquire an essential role for $\Delta nrp1$ mutants in both genetic backgrounds. The height of the bar indicates the number of reads associated with that insertion location.

depleted, 2) the phospholipid biosynthesis could be mainly regulated by a different gene, like *INO4*[28]. Precise mechanisms by which this reorganization is regulated remain to be seen. *MEC1* encodes an essential phosphoinositide (PI)-3-kinase-related protein kinase. This type of kinase

phosphorylates phosphoinositides on the 3-hydroxyl group of the inositol ring that produces secondary messengers in cell signaling [31]. It is a genome integrity checkpoint (specifically a mitosis entry checkpoint) responsible for DNA replication, repair, and telomere maintenance [32–35]. Furthermore, it regulates P-body formation under replication stress [36]. The transposon insertion map in the wild-type background shows two main regions void of transposons (panel C fig. 4.6); the most downstream region colocalizes with the kinase domain of this protein, presumably the domain that carries its essential function. Interestingly, for the $\Delta nrp1$ background, the kinase domain is fully enriched by transposons. Thus, we envision that the kinase activity of Mec1 is bypassed in $\Delta nrp1$ mutants, perhaps by buffering the signaling messengers that are products of this type of kinase activity. *TFB4* encodes a zinc ion binding protein involved in nucleotide excision repair and contributes to RNA polymerase II general transcription initiation factor activity [37, 38]. Analogously, $\Delta nrp1$ mutants show a moderate increase of transposons insertion along the gene. Similar to the essential gene *ARB1*, which is an ATP-binding protein involved in ribosome biogenesis. Especially, the transposon enrichment is more noticeable in the upstream side of the gene.

In general terms, these results imply that *NRP1* may regulate processes directly related to the inositol biosynthesis pathways, which affects important processes like DNA repair and replication. In addition, it potentially more distantly regulates transcription initiation and ribosome biogenesis, as previous studies related to *NRP1* have shown [12, 39].

Alternatively, to identify genes that become indispensable upon *NRP1* removal, we chose genes exhibiting significant negative genetic interactions with *NRP1*, and whose fitness values fall below the essential threshold for the mutant background but above the threshold for the wild-type background. We harbor four main candidate genes that meet these criteria, see bottom right square panel B fig. 4.6: *RPS0B*, *YLR198C*, *YER133W* and *CBS2*. The first one is a subunit of the cytosolic small ribosomal subunit (40S) [40], and it is involved in rRNA processing and export from the nucleus [41]. *YLR198C* is a dubious reading frame, which is unlikely to encode a functional protein [17]. *YER133W*, also known as *GLC7*, is involved in nuclear protein phosphatase involved in cell cycle regulation, DNA checkpoint signaling, and regulation of glycogen metabolism [42, 43]; localizes to spindle pole body, nucleolus, kinetochore, cell division site, and bud neck [44]. The last one, *CBS2*, is a mitochondrial translation regulator [45, 46].

Taken together, these results on the potential essential genes in $\Delta nrp1$ background suggest that *NRP1* is involved in the regulation of pathways that are complementary to some parts of the cell cycle regulation, RNA processing, and mitochondrial translation. Next section will dive into which submodules from the cell polarity function are more closely related to *NRP1* through its genetic genetic interactors.

4.2.7. Genes from the timing and the reaction-diffusion submodules of cell polarity show significant genetic interactions with *NRP1*

Overall, we have dissected some of the general functions that *NRP1* may be related to according to their predicted genetic interactors from transposon sequencing data. Now, we would like to take the opposite approach: look at a specific functional module and investigate which significant genetic interactors are assigned to it such that the possible linkage of *NRP1* with the target module is identified.

In this case, we select the cell polarity function, specifically the establishment of cell polarity module, due to its importance for the evolutionary experiment carried out by [6]. To conduct this

analysis, we carefully curated a subset of genes known to be part of the polarity establishment module based on a comprehensive literature review presented in [5]. The authors of that study dissected the polarity establishment module into five sub-modules: timing, mating, reaction-diffusion, actin, and bud scar-related functions, which are depicted in fig. 4.1.

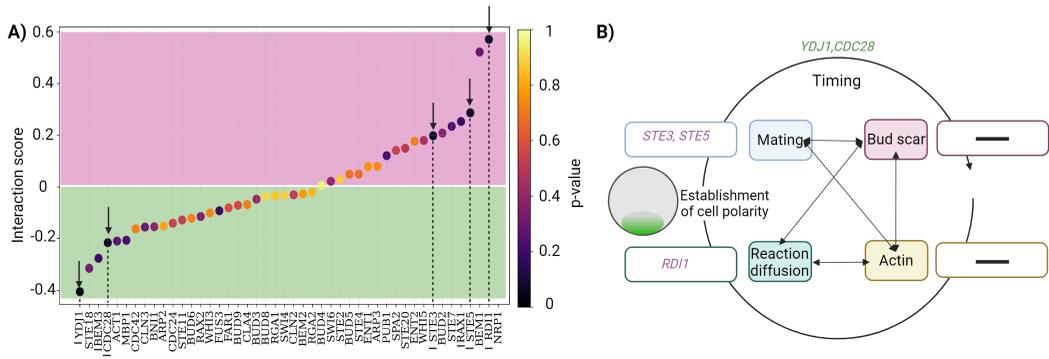


Figure 4.7. Polarity associated genes shows strong interactions with *NRP1*. **A)** Polarity establishment related genes, in the x-axis and their interaction score with *NRP1* in the y-axis. Darker dots represent statistically significant interactors. The arrows on top represent the genes with the lowest p-values. **B)** Pathways that lead to polarity establishment from [5]. Selected genes from **A)** are shown in the submodule they belong to. Green and purple colors correspond to negative and positive predicted genetic interaction with *NRP1*.

We sorted all the 46 polarity-associated genes according to their interaction score with *NRP1* and highlighted its significance through the color of the point, as panel A) from fig. 4.7 shows. The polarity genes cover a range of interaction scores from -0.4 to 0.6, and five have a p-value below 0.1. These genes are: *RDI1*, *STE3*, *STE5*, *YDJ1* and *CDC28*. The first three exhibit a positive interaction score, and the last two are negative. These genes cover the sub-modules of the reaction-diffusion (*RDI1*), the mating (*STE3*, *STE5*), and the timing (*YDJ1*, *CDC28*) pathways, panel B) fig. 4.7.

In summary, in this section, we adopt a different approach by focusing on the establishment of cell polarity, a crucial module in yeast biology. This module comprises five sub-modules: timing, mating, reaction-diffusion, actin, and bud scar-related functions. Analyzing the interaction scores between *NRP1* and the 46 genes associated with cell polarity, we identified five genes with significant scores: *RDI1*, *STE3*, *STE5*, *YDJ1*, and *CDC28*. These genes play roles in the reaction-diffusion, mating, and timing pathways.

Next, we integrate these findings concerning the possible linkage of *NRP1* with the polarity establishment module to elucidate suitable mechanisms for the recovery of polarity that is induced by knocking out *NRP1* in $\Delta bem1$ mutants.

4.2.8. Why does *NRP1* deletion rescue $\Delta bem1$ mutants?: Two plausible mechanisms

NRP1 deletion likely enhances the regulation of cell size at START in $\Delta bem1$ mutants.

The Cdc42 cell-polarization machinery of budding yeast is a robust module that, following the knockout of Bem1, a key player in the Cdc42-based polarity, cells regain their ability to polarize and divide by loss of three different genes [6], one of them is *NRP1*. However, the precise mechanisms of the polarity recovery remain unclear.

Based on the findings from the previous section about the strongest polarity interactors with *NRP1*, we identified two main submodules that *NRP1* deletion may be enhancing to recover polar-

ity in $\Delta bem1$ mutants: 1) Timing and 2) Reaction-diffusion submodules. We reasonably discarded the mating submodule to conceive the recovery mechanisms since this function does not play a role for the particular experiment conducted in [6], given that the yeast population consisted of haploid cells and, therefore, the chance to mate was excluded.

The first submodule to consider, the timing one, is represented by the strong novel negative and statistically significant interaction scores of *NRP1* with the genes *CDC28* and *YDJ1*, see panel A) from fig. 4.7. *CDC28* is a critical protein kinase that controls the timing of mitotic commitment, bud initiation, DNA replication, spindle formation, and chromosome separation[47]. *YDJ1* is a chaperone that recognizes cytosolic misfolded proteins during heat shock[48], and, more recently, it has been observed to play a critical role in determining cell size at START as a function of growth rate [49].

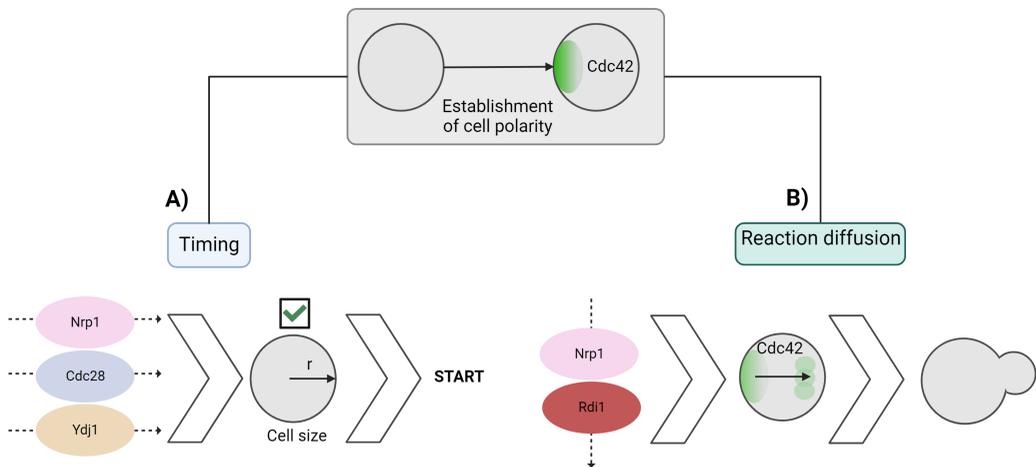


Figure 4.8. *NRP1* deletion may contribute to the recovery of polarity in $\Delta bem1$ mutants through the timing and the reaction diffusion pathways. **A)** From their predicted significant genetic interaction, we suggest that *Cdc28*, *Ydj1* and *Nrp1* may function together through complementary pathways in which that contribute to the proper timing of the cells at the start of the cell cycle. **B)** Following their predicted significant positive interaction, *Rdi1* and *Nrp1* may function together in the reaction-diffusion pathway, specifically in the *Cdc42* exchange rate from the membrane to the cytosol.

On the other hand, negative interactors are genes whose native expression and, therefore, function in the cell are required for $\Delta nrp1$ mutants. These gene pairs often involve parallel or alternative pathways contributing to a crucial cellular function, panel A) fig. 4.8. They could also belong to the same pathway but play critical roles. Additionally, these genes might be part of an essential protein complex where deleting both proteins results in an impaired essential protein, leading to cell fitness defects or cell death.

In this particular case, *CDC28* and *YDJ1* are known synthetic lethal[50], as well; thus, they both play a role in complementary pathways that contribute to regulating the initiation of the cell cycle.

Since *YDJ1* is the most extreme negative polarity interaction for *NRP1*, we hypothesize that *NRP1* is involved in a parallel pathway that contributes to cell size regulation at START. This function is especially critical in $\Delta bem1$ mutants, where yeast cells cannot initiate the bud formation, yet they undergo nuclear division without cellular division [51]. We predict that the polarity

restoration in $\Delta bem1\Delta nrp1$ mutations occurs through an improvement in the regulation of the cell size at the early stages of the cell cycle. Interestingly, *CDC28* and *YDJ1* are required genes for cell viability of $\Delta bem1$ mutants. On the contrary, Nrp1 native expression is toxic for $\Delta bem1$ mutants. Thus, Nrp1 function could hamper the proper regulation of cell size to initiate bud formation and progress in the cell cycle when polarity establishment can not rely on the function of *BEM1*.

Precise molecular mechanisms and right experimental tests are required to test this hypothesis. For example, firstly, we should test that in $\Delta nrp1$ mutants, a knockout in *YDJ1* and *CDC28* is harmful to the cells. Secondly, we can test how the dynamics of the different proteins during the early stages of the cell cycle correlate in the presence and absence of *BEM1*.

NRP1 deletion possibly enhances the Cdc42 rate exchange from the membrane to the cytosol in $\Delta bem1$ mutants.

The strongly predicted novel positive interaction between *RDI1* and *NRP1* is particularly interesting. Rdi1 is a Rho-GDP dissociation inhibitor that plays a crucial role in regulating the localization of Cdc42, a key protein involved in polarity establishment [52]. In the polarity establishment function, Rdi1 binds to membrane-bound Cdc42-GDP and transports it to the cytosol. Mutants lacking Rdi1 exhibit a slower exchange of Cdc42 from the membrane to the cytosol than wild-type cells, but similar levels of Cdc42 are eventually reached in these mutants [52]. In addition, $\Delta bem1$ mutants exhibit a reduced efficiency of Cdc42 condensation at one location in the plasma membrane, leading to the loss of efficient polarization and the possibility of multiple budding sites [7].

Interestingly, high-throughput studies [53] have shown that full knockout mutants of *bem1* require Rdi1 to proliferate, indicating a significant dependency on Rdi1 for cell survival and growth in the absence of Bem1, suggesting a potential compensatory mechanism involving Rdi1 in *bem1* mutants.

Based on these findings, we speculate that Nrp1 deletion in *bem1* mutants could enhance the rate of Cdc42 exchange from the membrane to the cytosol through its positive interaction with Rdi1. As a result, *bem1nrp1* mutants may exhibit better polarization because they can maintain a highly dynamic flow of Cdc42 in and out of the membrane [52].

To investigate this effect further, follow-up experiments such as FRAP (Fluorescence recovery after photobleaching) experiments on the formed buds of $\Delta bem1\Delta nrp1$ and $\Delta bem1$ mutants can provide valuable insights into the dynamics of Cdc42 and its role in the enhanced polarization observed in the double knockout mutants.

In conclusion, Nrp1 deletion may rescue $\Delta bem1$ mutants by enhancing the cell size regulation before bud initiation and by increasing the dynamic flow of *CDC42* out of the polarity spot, facilitating the progress of further stages in the cell cycle. However, it is important to note that the predicted interactions and proposed mechanisms are based on computational analysis and literature-based predictions. Further experiments, such as growth rate analysis, time-lapse fluorescence microscopy, and FRAP, are required to precisely characterize these mechanisms and validate the role of *NRP1* deletion in rescuing *bem1* mutants.

4.3. Discussion

This chapter has developed a methodology to study an example of a weak-effect gene. Specifically, we focus on the gene *NRP1*, which is currently annotated as having an unknown function

in budding yeast; however, it may play a crucial role in cell survival in cells harboring a complete gene knockout in *BEM1*.

Our study aimed to elucidate the functional role of *NRP1* using diverse experimental approaches. Initially, we investigated the potential impact of Nrp1's RNA binding domain on gene expression. However, our analysis revealed no significant alterations in the expression of approximately 1200 proteins, suggesting that *NRP1* does not serve as a regulator of gene expression.

Subsequently, we performed a comprehensive genome-wide analysis of *NRP1* genetic interactors to infer its function through interactions with other genes. The functional enrichment analysis of negative genetic interactors identified via SATAY data indicated potential involvement of *NRP1* in processes related to determining critical cell size at the START point of the cell cycle, endocytosis, prion formation, and fermentation growth (panel F, fig. 4.5). Additionally, enrichment analysis of positive interactors provided insights into pathways potentially associated with *NRP1*, including exit from G0 during the cell cycle, spindle morphology, and respiratory growth.

Remarkably, a recent study, [54], has shown the emerging role of RNA binding domain proteins in mitotic spindle formation in eukaryotes. Treatment with transcription inhibitors or RNase disrupts the mitotic spindle structure independent of active translation [55, 56], implying that functional mRNAs or non-coding RNAs (ncRNAs) act as regulators or structural components of the mitotic spindle, from [57]. In addition, the fission yeast ortholog of Nrp1, known as Dri1, has been directly linked to the assembly of the mitotic spindle [57]. Studies on fission yeast have shown that the absence of Dri1 results in a significant decrease in the amount of kinesin-14/Klp2 protein on the spindle, and this mechanism can rescue *cut7* mutants. Dri1 in fission yeast also contains an RNA binding domain involved in heterochromatin assembly.

To explore whether Nrp1 deletion could rescue $\Delta kip1$ or $\Delta cin8$ mutants, which are the orthologous genes to *CUT7* in budding yeast, we look to whether those genes exhibit a significant interaction with *NRP1* from our dataset. However, we found no significant interaction for those genes with *NRP1*, as their interaction score values were close to zero (0.02 for *KIP1* and -0.09 for *CIN8*). Additionally, we investigated whether Kar3, the orthologous protein to kinesin-14/Klp2 in budding yeast, might genetically interact with *NRP1*. We observed a mild, not significant⁵, positive interaction between Kar3 and Nrp1 (interaction score $\epsilon=0.14$, p-value=0.3), suggesting a far functional relationship between these proteins. Further studies are required to elucidate the specific role of this interaction and its implications in spindle assembly and other cellular processes.

The other part of this study focuses on elucidating potential novel interaction partners with the polarity establishment module to explain the polarity recovery in $\Delta bem1 \Delta nrp1$ mutants. We quantified the predicted genetic interactions of *NRP1* with all the genes assigned to this module [5] and ranked them based on their statistical significance. We identified that the most significant connection of *NRP1* with polarity establishment is through the timing and reaction-diffusion pathways, specifically via its significant genetic interactions with *CDC28*, *YDJ1* and *RD11*. We claim that *NRP1* deletion can enhance regulation of the critical cell size before bud formation and enhance the dynamic flow of *CDC42* from the membrane to the cytosol enabling the progress to further cell cycle stages in $\Delta bem1$ mutants.

On the same line, [5] proposes that the function of *NRP1* may be linked to the function of the protein *WHI3*, as they share a considerable sequence similarity. *WHI3* is known to localize to stress granules and indirectly regulate the G1/S phase transition via *CLN3*, modulating the stability and translation efficiency of *CLN3* [58]. Additionally, *WHI3* has been identified as a genetic suppressor

⁵For this specific dataset

for $\Delta bem1$ cells [53].

From our data analysis results, we find that neither *WHI3* nor *CLN3* exhibit a significant genetic interaction with *NRP1*. However, we identified genetic interaction partners of *NRP1* related to prion formation and cell cycle regulation at the G1 phase. In addition, we found that *CDC28* is a strong negative interactor of *NRP1*, which associates with *CLN3* to setting the size threshold at which cells pass through START (commitment to duplication and division) [59].

We lastly propose that the conservation of Nrp1 across the Ascomycota kingdom [8] could be attributed to its potential involvement in the timing regulation throughout the first stages of the cell cycle. This hypothesis can expand on the claims by [51] that proposes that distinct functional mechanisms contribute to cell polarization in budding yeast under different regimes of protein copy numbers. Thus, adding the cell size control at the early stages of the cell cycle could be another functional mechanism that regulates *CDC42* based cell polarization in budding yeast.

4

4.4. Outlook and future experiments

The proposed experimental checks are crucial for confirming or discarding the suggested hypotheses for the function of Nrp1. First, we must verify that predicted essential genes in $\Delta nrp1$ mutants lead to a severe fitness defect.

We propose the construction of $\Delta nrp1\Delta rps0b$ and $\Delta nrp1\Delta cbs2$ strains: This involves creating double knockout strains of *NRP1* with the genes *RPS0B* and *CBS2*. If the resulting double knockout strains exhibit non-viability or severe fitness defects, it would confirm that these genes are synthetic lethal, or aggravating interaction partners for *NRP1*. Viability can be assessed through spot plating and growth rate population experiments.

Secondly, the evaluation of essential gene dispensability in *nrp1* mutants: This entails examining whether certain essential genes in *wild type* become dispensable in *nrp1* mutants. Double knockout strains such as $\Delta nrp1\Delta mec1$ and $\Delta nrp1\Delta ino2$ can be constructed. If these strains exhibit an increase in growth rate or colony size compared to their *wild type* counterparts, it would indicate that the essential genes have become dispensable in the presence of *NRP1* deletion. To ensure the specificity of the results, the essential gene should be removed in the $\Delta nrp1$ mutant background to avoid suppressor mutations or aneuploidies caused by the lethality of those genotypes in *wild type* strains.

After confirming our predictions from the SATAY data analysis, we plan to delve deeper into dissecting the function of *NRP1*. This chapter proposes that *NRP1* is involved in the polarity establishment function primarily by enhancing the regulation of cell size at START, by *YDJ1* and the *CDC42* exchange rate from the membrane to the cytosol through its interaction with *RD11*. To test those suggested functions of *NRP1* in $\Delta bem1$ mutants, we recommend:

Polarization enhancement in *bem1nrp1* versus *bem1* mutants: To precisely quantify the dynamics of polarity establishment in *bem1nrp1* mutants compared to *bem1* mutants, we recommend conducting fluorescence microscopy experiments focused on studying the Cdc42-based polarity. This approach will provide valuable insights into the role of *NRP1* deletion in enhancing polarity and recovering $\Delta bem1$ mutants' fitness. By visualizing and analyzing the spatial and temporal behavior of *CDC42*, we can better understand how *NRP1* influences the polarization process during yeast cell budding.

Timing pathway via *YDJ1* and *CDC28*: We suggest to perform time-lapse fluorescence microscopy, tagging *YDJ1*, *CDC28* and *NRP1* with different fluorescent markers. In addition, tagging *WHI5* as a marker to indicate the START of the cell cycle [60] would be beneficial to have a reference on the cell cycle stage during the experiment. We propose to arrest all cells in G1 to synchronize them in the same cell cycle point and measure their dynamic correlation, especially after START until bud formation. This would be measured in mutants harboring *BEM1* and *NRP1* knockouts, having wild-type cells as control. Cell radius can be quantified in the different mutants and correlated with their temporal dynamic throughout the different cell cycle stages.

Reaction diffusion pathway via *Rdi1*: Performing fluorescence recovery after photobleaching (FRAP) experiments of Cdc42-mNeonGreen in both $\Delta bem1$ and $\Delta bem1\Delta nrp1$ mutants will allow us to examine the exchange of Cdc42 from the plasma membrane to the cytosol. To ensure synchronized cell cycles and photobleaching at the nascent cap, the mutants should be arrested in the G1 phase. This analysis will provide information on the dynamics of Cdc42 and how *Nrp1* deletion affects its exchange rate. Additionally, we recommend exploring whether *NRP1* deletion can rescue the fitness defects caused by *RD11* knockout in $\Delta bem1$ mutants. This experiment aims to assess whether the positive interaction between *RD11* and *NRP1* persists after *BEM1* deletion.

4.5. Materials and Methods

Materials

List of strains, plasmids, and primers

#	Name	Genotype	Source
1	yES1a,b	<i>MATx can1-100,leu2-3,112,his3-11,15,ura3Δ, BUD4-S288C</i>	Laan Lab collection
2	yES11a,b	<i>MATx can1-100,leu2-3,112,his3-11,15,nrp1::URA3, BUD4-S288C</i>	This study
3	yES32a,b	<i>MATx can1-100,leu2-3,112,his3-11,15,ura3Δ, bem1::pGal-BEM1 BUD4-S288C</i>	This study
4	yES39a,b	<i>MATx can1-100,leu2-3,112,his3-11,15,bem1::pGal-BEM1, nrp1::URA3, BUD4-S288C</i>	This study
5	yES40a,b	<i>MATx can1-100,leu2-3,112,his3-11,15, bem1::pGal-BEM1 BUD4-S288C</i>	This study
6	yLIC133a	<i>MATx can1-100,leu2-3,112, his3-11,15, ura3Δ,ade2Δ BUD4-S288C</i>	This study
7	yLIC136a	<i>MATx can1-100,leu2-3,112, his3-11,15, ura3Δ,ade2Δ nrp1::HYGRO, BUD4-S288C</i>	This study
8	yWT03	<i>MATx can1-100,leu2-3,112, his3-11,15, ura3Δ,ade2Δ ho::osTIR1-3xMyc-kanMX, bem1::mCherry-AID, BUD4-S288C</i>	This study
9	yLL137	<i>MATx can1-100,leu2-3,112, his3-11,15, ura3Δ nrp1::HYGRO BUD4-S288C</i>	This study

Table 4.1. List of strains used in this chapter

#	Name	Source	Description
1	pES001	Laan Lab	<i>NRP1</i> flanking
2	pES012	This study	Gal1p- <i>BEM1</i>
3	pES013	This study	<i>NRP1-URA3</i>
4	pBK549	[61]	MiniDs Transposons and Gal1-Transposase
5	pOsTIR1 w/o GFP	Addgene catalog # 102833	OsTIR1 insertion
6	PG23A	Addgene catalog # 102884	AID TAG
7	pLL112	Laan lab	<i>URA3</i> fragment

Table 4.2. List of plasmids used in this chapter.

Name	Sequence	Source
oES49	GTATCTGGAGATGAGAAGGA TTGGGGATCCGGTTTGAGA AAGCTACTCCTC	This study
oES53	GAGGAGTAGCTTTCT CAAACCGGATCCCCAATCCTTCTC ATCTCCAGATAC	This study
oES57	TCGGTAATGGTTTTAATAG TTCAATACGTTGGTAGc gcggtggagctccaatt	This study
oES62	AATTGTTTCATGTACAAA ATCTTAGAT TTCTCTTGCAAAAA cataATTTCTGTATCAAAAA TCACAATAGCAAT GCACAAT	This study
oES77	GAGTAGGCGTTCATAAC TTTGTGTATCT	This study
oES82	CAAGCCTTGAAGATTT GCATAACGG	This study
oES83	CGATGACGAAGACGATGAAGACA	This study
oES84	CTGCCGCTGGTGAAGAAATTC	This study
oES28/SPY1	TGCCGTGGAATCAT ACGAACCG	This study
oES29/SPY2	GGTAGAGTGTGGAA AGGCTCTGC	This study
oES58/SPY3	GTGCCGCTGCIT TCGAG	This study

Name	Sequence	Source
oES60/Spy5	GGAACAAGGCC GGCC	This study
Seq4_Fwd	AATTATCCTGGG CACGAG	Wessel Teunisse Master Thesis
Seq4_Rev	ACTGTAAGATTC CGCCAC	Wessel Teunisse Master Thesis
Bem1-mCherry-AID-FW	ACCTAGTGAATCTTCCCGATT ATATATCTCGCTC	Wessel Teunisse Master Thesis
Bem1-mCherry-AID-RV	AGGGAGCCACATTATC CTTTGACACATATG	Wessel Teunisse Master Thesis
olic15	GCTGTGGtatgggctactctcCTTGATTG TTTTGTCCGATTTT CTTGTTTTCTTG	This study
olic16	CAAGAAAAACAAGAAAAATCGG ACAAAACAATCAAG gagagtgaccataCCACAGC GATGTAATCATAACAAA	This study
olic17	GCCTAAAA AATAGGTATATC GTGAGTTTAGTATA CATGCATTTACTATAATACAG	This study
olic18	CTGTATTATAAGTAAATGCA TGTATACTAAA CTCAGGATATACCTATT TTTTAGGCTTTGTTATGATT ACATC	This study

Name	Sequence	Source
olic20	GATGTAATCATAACAAA	This study
	GCCTAAAA AATAGGTATATCCTTG ATTGTTTTGTCGATTTTCTT GTTTTTCTTG	
olic21	CAAGAAAAACAAGA	This study
	AAATCGGAC AAAAACAATCAAGGAT ATACCTATTTTTAGGCT TTGTTA TGATTACATC	
olic24	ATTACAGCTATGCTG	This study
	ACAAATGACTCTTG	
olic26	GCTATCCTCGGTTC	This study
	TGCATTGAGC	

Table 4.3. List of primers used in this chapter.

Yeast strains construction methods

***nrp1* full knockout mutants.** The *URA3* gene was PCR amplified from the plasmid pES001 with primers oES77 and oES82. The amplified DNA fragment was transformed in yES1 yeast cells using the lithium acetate high-efficiency yeast transformation protocol [dx.doi.org/10.17504/protocols.io.gzrbx56](https://doi.org/10.17504/protocols.io.gzrbx56), yielding the strain yES11. The inserted fragment was checked by Sanger sequencing (Macrogen <https://dna.macrogen.com/>) of the PCR fragment from yES32 genomic DNA and primers oES28/oES29. Strains were stored at -80°C at 80% (w/v) glycerol.

Galactose inducible *BEM1* expression. The native *BEM1* promoter was replaced by the galactose promoter by transforming the amplified DNA fragment from the plasmid pES012 with primers oES49 and oES53, into the backbone of the strain yES01, giving rise to the strain yES32. The inserted fragment was checked by Sanger sequencing (Macrogen <https://dna.macrogen.com/>) of the PCR fragment from yES32 genomic DNA and primers oES58/oES60. Strains were stored at -80°C at 80% (w/v) glycerol.

Knockout of *URA3* and *ADE2* for SATAY. We kick out the *ADE2* locus from the yLL3a strain following a two step procedure. First, the *URA3* genes were inserted in *ADE2* locus and then transformed by an empty sequence. The procedure consisted in making two DNA fragments (OEP1 and OEP2) that were transformed sequentially in yLL3a, using the lithium acetate high-efficiency yeast transformation protocol [dx.doi.org/10.17504/protocols.io.gzrbx56](https://doi.org/10.17504/protocols.io.gzrbx56). The DNA fragments were produced by the method of overlap extension PCR; see [dx.doi.org/10.17504/protocols.io.psnndnd](https://doi.org/10.17504/protocols.io.psnndnd) for a whole description of this method. The OEP1 is the DNA fragment used to insert the *URA3* gene in the *ADE2* locus, and the OEP2 is the DNA fragment used to kick out the *URA3* gene from this location. For the OEP1, we need to produce three DNA fragments that will be glued together through PCR. The first DNA fragment corresponds to the upstream region of the *ADE2* gene amplified by PCR. Specifically, for this PCR reaction, the template was genomic DNA from yLL3a and primers olic24/olic15 (I), the second to the *URA3* gene fragment amplified from the plasmid pLL112 with primers olic16 and olic17 (II). Lastly, the downstream region of *ADE2*, which we obtained by PCR with the yLL3a genomic DNA and olic18/olic26 primers (III). Then, all three DNA fragments (I+II+III) were taken as templates for the overlap extension PCR method with primers olic24/olic26. The OEP2 fragment glued two DNA fragments with no homology to the *URA* gene. Firstly, the fragment from the yLL3a genomic DNA and primers olic24/olic20 and, secondly, the fragment from the same template and primers olic21/olic26. We checked the expected length of OEP1 and OEP2 by DNA electrophoresis and sequenced by Sanger sequencing with Macrogen. This method produces the strain yLIC133. Further, the transformation of yLIC133 with an amplified DNA fragment from the yLL137 strain with primers oES83/oES84 yields the strain yLIC136.

Auxin inducible degron system on *BEM1*. The SATAY library of the $\Delta bem1$ mutant was generated using an auxin inducible degron system [62]. This system will deplete Bem1 from the cell upon the presence of auxin. Its application is to avoid as much as possible the presence of suppressors that quickly sweep the $\Delta bem1$ population toward higher fitness values. This method can conditionally induce the degradation of any protein by the proteasome upon the addition of the auxin hormone, see Panel **A**) from fig. 4.9. This system requires two components: the AID tag that tags the protein of interest for degradation and OsTIR1, which is a requisite for ubiquitination of the AID tag in the presence of auxin. The OsTIR1-3XMyC gene fragment was obtained from the plasmid pOsTIR1 w/o GFP (Addgene catalog # 102883) through PCR amplification with primers Seq4_Fwd and Seq4_Rev, shown in table 4.3. Next, the amplified DNA fragment was integrated

into the *HO* locus of strain yWT01a (taken from a colony restreak of yLIC133a) via homologous recombination, which resulted in strain yWT02a. For the construction of the gene fragment that enables the conditional depletion of *BEM1*, the plasmid pG23A (Addgene catalog # 102884) was modified using Gibson assembly to replace the sequences coding for the *CDC14* N-terminus and the downstream flanking site with those of *BEM1*. Further, the *BEM1-mCherry-AID* DNA fragment was amplified by PCR with primers Bem1-mCherry-AID-FW and Bem1-mCherry-AID-RV and integrated into strain yWT02a, yielding yWT03a. The transformed *BEM1-mCherry-AID* in yWT03a was verified with Sanger sequencing (Macrogen), which revealed a single SNP (A to T conversion) at position 1433 of the coding sequence of *BEM1*. This variant occurs naturally in the S288C genomic background of the *S. cerevisiae* was therefore assumed to have no detrimental effects on the function of Bem1. Strains were stored at -80°C at 40% (w/v) glycerol. Plasmids pOsTIR1 w/o GFP Addgene catalog # 102883) and pG23A (Addgene catalog # 102884) were a gift from Matthias Heinemann <https://www.rug.nl/staff/m.heinemann/cv>. Sequencing of the *bem1-aid* gives rise to the enrichment of transposon insertions and reads in the AID, which will result in the highly beneficial effect of reconstituting Bem1 back in the cell, see panel **B**) from fig. 4.9. A considerable number of transposons insertions map to genes like *ADE2* or *BEM1*; thus, they were discarded for the fitness analysis of this library since they are enriched in non-biological. Mutations in the AID system will render a full copy of Bem1, which is highly beneficial in this population. Transposons from the *ADE2* gene will correspond to *ADE2* gene fragment from the transformation plasmid.

4

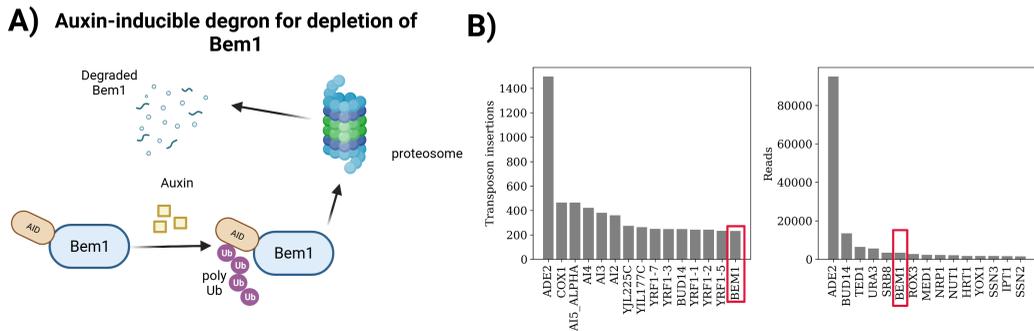


Figure 4.9. Auxin inducible degron system

SATAY libraries.

Media

Standard yeast culturing. Standard culturing was performed in YPD (10g/L Yeast extract, 20g/L Peptone, 20g/L dextrose), SC (6.9 g/L Yeast nitrogen base, 0.75 g/L Complete supplement mixture, 20g/L dextrose). All liquid media were filtered and sterilized to prevent degradation of the media components, especially amino acids.

Non-standard culturing media. For *ade⁻* strains, the growth media was supplemented with 22mg adenine per 400 mL batch of media during the SATAY procedure. For culturing the *bem1-aid* strains, 0.1M stock solutions of indole-3-acetic acid (IAA) were prepared by dissolving 175mg IAA in 2mL of 100% absolute ethanol. The final volume was adjusted to 10ml, adding extra ethanol when necessary to prevent precipitation of IAA. The stock solution was filter sterilized

Strain	# of transposons	# of reads	# of transposons in CDS	# of reads in CDS
yLIC133_a	423524	23428274	194719	7246573
yLIC133_b	414347	24441674	189469	7577024
yLIC136_a	467104	28569602	287127	16494324
yLIC136_b	510327	25383712	301283	16057911

Table 4.4. Characteristics of the SATAY libraries of the wild type and *nrp1* mutant yeast strains, after the transposon sequencing data analysis. From left to right in the columns, the strain name, the number of transposons mapped, the number of reads mapped, the number of transposons mapped to coding sequences(CDS), and the number of reads mapped to CDS.

using a 0.2 μm syringe filter, aliquoted, and stored at -20°C .

Methods

Shot-gun mass spectrometry

The protocol employed for the shotgun mass spectrometry assay is the same as the one followed by the publication [63]: *den Ridder, Maxime, et al. "Proteome dynamics during the transition from exponential to stationary phase under aerobic and anaerobic conditions in yeast." Molecular and Cellular Proteomics 22.6 (2023).*

Yeast cell lysis, protein extraction, and proteolytic digestion. Cell pellets were resuspended in a 100 mM Triethylammonium bicarbonate (TEAB) lysis buffer containing 1% SDS and phosphatase/protease inhibitors. Yeast cells were lysed by glass bead milling by 10 cycles of 1-minute shaking alternated with 1 min rest on ice. Proteins were reduced by adding 5 mM DTT and incubating for 1 hour at 37°C . Subsequently, the proteins were alkylated for 60 min at room temperature in the dark by adding 50 mM acrylamide. Protein precipitation was performed by the addition of four volumes of ice-cold acetone (-20°C), followed by 1 hour freezing at -20°C . The proteins were solubilized using 100 mM ammonium bicarbonate. Proteolytic digestion was performed by Trypsin (Promega, Madison, WI), 1:100 enzyme to protein ratio, and incubated at 37°C overnight. Solid phase extraction was performed with an Oasis HLB 96-well $\mu\text{Elution}$ plate (Waters, Milford, USA) to desalt the mixture. Eluates were dried using a SpeedVac vacuum concentrator at 50°C and frozen at -80°C .

Quantitative temporal proteome analysis. Desalted peptides were reconstituted in 100 mM TEAB, and TMT10-plex reagents (Thermo) were added from stocks dissolved in 100% anhydrous acetonitrile (ACN). Peptides were mixed with labels in a 1:8 ratio (12.5 μg to 100 μg) and incubated for 1 hour at 25°C and 400 rpm and the labeling reaction was stopped by the addition of 5% hydroxylamine to a final concentration of 0.4%. Labeled peptides were then mixed in at approximating equal quantities. Two bridging samples were included in each TMT10-plex experiment to improve comparability between different experiments. The bridging sample was a mixture of the three biological replicates of MG yeast under aerobic conditions in the mid-stationary phase. Peptide solutions were diluted with water to obtain final acetonitrile (ACN) concentration lower than 5%. Solid phase extraction was performed to desalt the final peptide mixture. Desalted peptides were frozen at -80°C for 1 hour and dried by vacuum centrifugation. Peptides were finally resuspended in 3% ACN/0.01% TFA prior to MS analysis to give an approximate concentration of

500 ng per μL .

Shotgun proteomic analysis. An aliquot corresponding to approximately 1 μg protein digest was analyzed using an one-dimensional shot-gun proteomics approach [64]. Briefly, the samples were analyzed using a nano-liquid-chromatography system consisting of an EASY nano-LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 μm x 150 mm, 2 μm , Cat. No. 164568), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). The flow rate was maintained at 350 nL/min over a linear gradient from 5% to 25% solvent B over 180 min, then from 25% to 55% over 60 min, followed by back equilibration to starting conditions. Data were acquired from 5 to 240 min. Solvent A was H_2O containing 0.1% formic acid (FA), and solvent B consisted of 80% ACN in H_2O and 0.1% FA. The Orbitrap was operated in data-dependent acquisition (DDA) mode acquiring peptide signals from 385–1250 m/z at 70 K resolution in full MS mode with a maximum ion injection time (IT) of 75 ms and an automatic gain control (AGC) target of 3E6. The top 10 precursors were selected for MS/MS analysis and subjected to fragmentation using higher-energy collisional dissociation (HCD). MS/MS scans were acquired at 35 K resolution with an AGC target of 1E5 and IT of 100 ms, 1.2 m/z isolation width, and normalized collision energy (NCE) of 32.

4

Mass spectrometry data analysis

Processing of mass spectrometric raw data. Data were analyzed against the proteome database from *Saccharomyces cerevisiae* (Uniprot, strain ATCC 204508 / S288C, Tax ID: 559292, July 2020) using PEAKS Studio X (Bioinformatics Solutions Inc., Waterloo, Canada) [16], allowing for 20 ppm parent ion and 0.02 m/z fragment ion mass error, 3 missed cleavages, acrylamide and TMT10 label as fixed and methionine oxidation and N/Q deamidation as variable modifications. Peptide spectrum matches were filtered against 1% false discovery rates (FDR) and identifications with ≥ 2 unique peptides. Changes in protein abundances between different time points using the TMT quantification option provided by the PEAKSQ software tool (Bioinformatics Solutions Inc., Canada). Auto normalization was used for the quantitative analysis of the proteins, in which the global ratio was calculated from the total intensity of all labels in all quantifiable peptides. Quantitative analysis was performed using protein identifications containing at least two unique peptides, in which peptide identifications were filtered against 1% FDR. The significance method for evaluating the observed abundance changes was set to ANOVA, and the significance score was expressed as the $-10 \times \log_{10}(p)$, where p is the significance testing p-value. The p-value represents the likelihood that the observed change is caused by random chance. Results from PEAKSQ were exported to 'proteins.csv' containing the quantified proteins.

Data availability: Mass spectrometric raw data have been deposited in the 4TURearchData repository under the doi : <https://doi.org/10.4121/7fcd8e37-d105-4a7a-a546-37d55b0a57dd.v1>

Population growth assays.

Cells were grown in 96 well plates (Thermo Fisher # 267427) until saturation for two days. The plates were filled with 2mL sterile MiliQ in each edge moat, and the outer 36 wells with 100 μl media as control wells to prevent positional bias during measurement. In the remaining wells, 5 μl of cells from glycerol stocks were inserted in 100 μl of media. The plates were placed in a shaker at a speed of 100rpm-150rpm in a 30°C room. For growth experiments, we used a factor of 200X of dilution for the saturated samples. Cells in the saturation plate will sediment. Therefore, before pipetting up (with a multi-pipette), we suggest stirring gently and scratching on the

bottom of the well plate or pipet up and down to resuspend them. In the plate reader (Biotek Epoch 2), we began the measurement with 1000s of shaking (amplitude 1 mm, linear) until the temperature reached 30°C. After the target temperature is met, each measurement cycle had five steps: 90s linear shaking with 2 mm amplitude, 90s orbital shaking with 1.5 mm amplitude, 90s linear shaking with 1 mm amplitude, 10s waiting, and Absorbance measurement at 600nm every 7 minutes over 48 hours. Doubling times were extracted from the growth curves using an in-house Matlab code.

Genetic interactions calculations

After computing the fitness for every gene deletion, we quantify genetic interactions. As well as genes that are in poor transposon-enriched flanking regions. Digenic interactions are computed using the function *digenic_GI.py* from https://github.com/SATAY-LL/comparative-analysis-among-strains/blob/main/src/functions_interaction_computations.py.

Volcano plots

Volcano plots are computed using the function *annotate_volcano.py* from https://github.com/SATAY-LL/comparative-analysis-among-strains/blob/main/src/annotate_volcano.py. Previous to use this function the user must classify every genetic interaction in "pos" (for positive), "neg" (for negative), and "none" (for no interaction). We use the function *classify_GI* from https://github.com/SATAY-LL/comparative-analysis-among-strains/blob/main/src/functions_interaction_computations.py to extract the positive and negative genetic interactors.

Enrichment analysis

Functional enrichment analysis with Gseapy. The enrichment analysis shown in panel C-F) from fig. 4.5 and fig. 4.10 were made using the Python library **Gseapy** <https://gseapy.readthedocs.io/en/latest/introduction.html>. A regular code snippet using this library looks like this:

Listing 4.1. Gseapy example code snippet

```
import gseapy as gp

goi=genes_of_interest
yeast = gp.get_library_name(organism='Yeast')

sets=[ 'GO_Biological_Process_AutoRIF',
'GO_Cellular_Component_AutoRIF',
, 'GO_Molecular_Function_AutoRIF',
'Pfam_Domains_2019', 'Phenotype_AutoRIF',
'WikiPathways_2018' ]

# ### enrichment

data=[]

for i in np.arange(0, len(sets)):

    enr=gp.enrichr( gene_list=goi,
```

```

gene_sets=sets[i],
organism='Yeast',
outdir='path_to_output_dir',
cutoff=0.1 # p_value
)
data.append(enr.res2d)

```

The *data* python array will contain in every dimension the enrichment according to the *sets* provided. Further bar plots are generated with the *barh* python function from *matplotlib* library of the sorted *Adjusted P-value* variable from any interesting dimension from *data* python array.

4

Appendix

Appendix: Enrichment analysis for the mass spectrometry proteins set

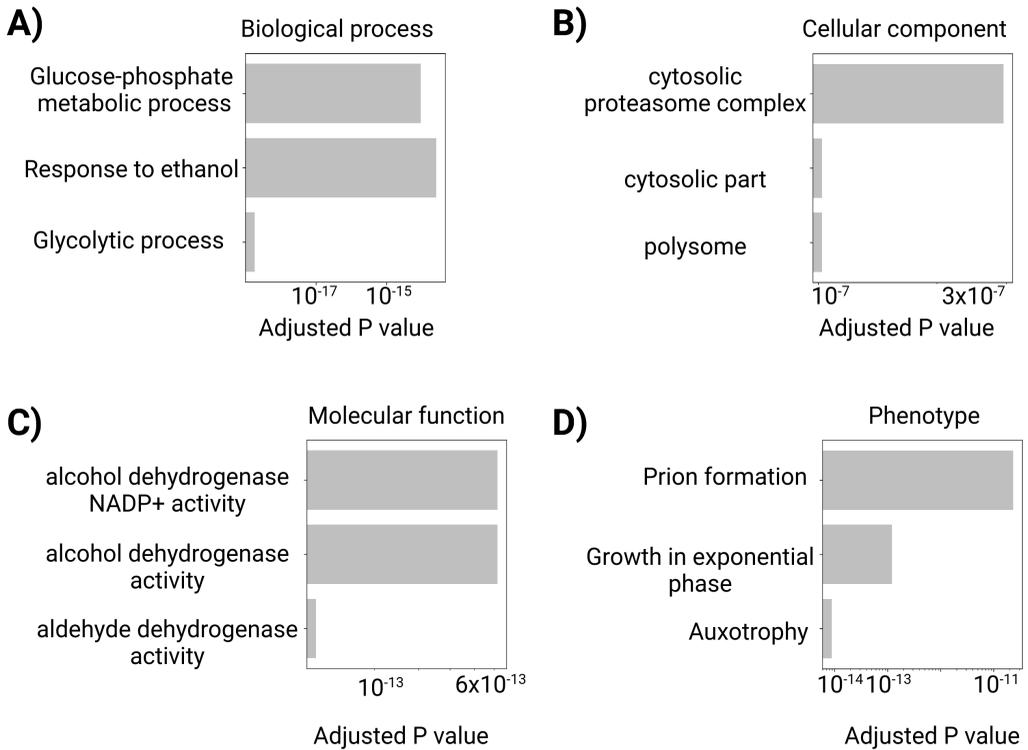


Figure 4.10. Enrichment analysis for significant proteins from shot-gun mass spectrometry technique.

SATAY readouts are validated with population growth rate measurements

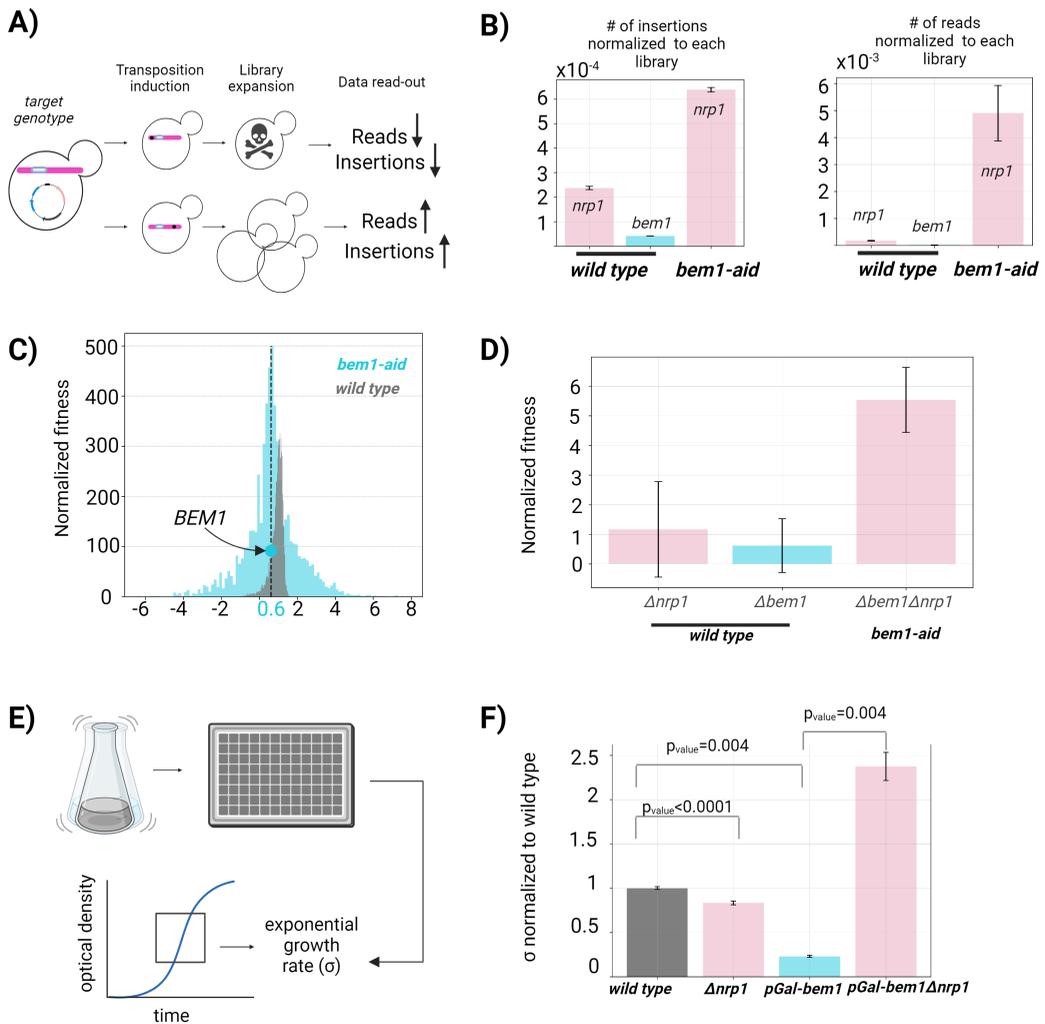


Figure 4.11. SATAY readouts are validated with population growth rate measurements. **A)** A simplified schematic of how the SATAY readouts relate to mutant abundance. We start with a specific monoclonal population when inducing transposition. Each transposon will jump to a random location in the genome. Two cases can show up, whether or not the disrupted gene is required for growth. If it is, this population will eventually vanish, and if it is not, we expect a proliferation of cells carrying the same transposon in the population. The transposon insertion and reads as the SATAY readouts indirectly correlate with mutant abundance. **B)** Insertion and read profiles of *NRP1* and *BEM1* in *wild type* and $\Delta bem1$ mutants. Error bars indicate the standard deviation of the insertion(read) profile across technical replicates (2). **C)** Genome wide normalized fitness distributions in *bem1-aid* and in *wild type* cells. In *bem1-aid* background, the fitness effects of the other genes are highly altered compared to the *wild type*. The distributions were normalized such that the distribution's median corresponds to the knockout mutant's growth rate. **D)** $\Delta nrp1$ knockout in $\Delta bem1$ background is predicted to rescue $\Delta bem1$ phenotype by the increase in its fitness value. The *nrp1* knockout in *wild type* does not show a clear difference in its mean fitness value. Error bars represent the variance of fitness values within the gene of interest. This variance is corrected by the number of insertions in that gene, considering them as different replicates with the same mutation. **E)** Population growth experimental setup. We put the different samples in a 96-well plate that goes into a plate reader to measure the optical density of every well at different time points for 24 hours. **F)** Exponential growth rates quantification from population growth experiments. Consistently, deletion of *nrp1* significantly rescues $\Delta bem1$ knockouts, increasing the fitness by approximately 5-fold. Likewise, *nrp1* mutants do not show a clear difference with *wild type*.

Bibliography

1. Park, H.-O. & Bi, E. Central roles of small GTPases in the development of cell polarity in yeast and beyond. *Microbiology and molecular biology reviews* **71**, 48–96 (2007).
2. Drubin, D. G. & Nelson, W. J. Origins of cell polarity. *Cell* **84**, 335–344 (1996).
3. Diepeveen, E. T., de la Cruz, L. I. & Laan, L. Evolutionary dynamics in the fungal polarization network, a mechanistic perspective. *Biophysical Reviews* **9**, 375–387 (2017).
4. Irazoqui, J. E. & Lew, D. J. Polarity establishment in yeast. *Journal of cell science* **117**, 2169–2171 (2004).
5. Daalman, W. K.-G., Sweep, E. & Laan, L. The Path towards Predicting Evolution as Illustrated in Yeast Cell Polarity. *Cells* **9**, 2534 (2020).
6. Laan, L., Koschwanez, J. H. & Murray, A. W. Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *Elife* **4**, e09638 (2015).
7. Brauns, F. *et al.* Adaptability and evolution of the cell polarization machinery in budding yeast. *bioRxiv*, 2020–09 (2020).
8. Diepeveen, E. T., Gehrman, T., Pourquié, V., Abeel, T. & Laan, L. Patterns of conservation and diversification in the fungal polarization network. *Genome Biology and Evolution* **10**, 1765–1782 (2018).
9. Reynaud, A., Facca, C., Sor, F. & Faye, G. Disruption and functional analysis of six ORFs of chromosome IV: YDL103c (QRI1), YDL105w (QRI2), YDL112w (TRM3), YDL113c, YDL116w (NUP84) and YDL167c (NRP1). *Yeast* **18**, 273–282 (2001).
10. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146–158 (2009).
11. Hogan, D. J. *et al.* Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS biology* **6**, e255 (2008).
12. Buchan, J. R., Muhlrud, D. & Parker, R. P bodies promote stress granule assembly in *Saccharomyces cerevisiae*. *The Journal of cell biology* **183**, 441–455 (2008).
13. Wade, C. H., Umbarger, M. A. & McAlear, M. A. The budding yeast rRNA and ribosome biosynthesis (RRB) regulon contains over 200 genes. *Yeast* **23**, 293–306 (2006).
14. Den Ridder, M., Daran-Lapujade, P. & Pabst, M. Proteome dynamics during transition from exponential to stationary phase under aerobic and anaerobic conditions in yeast. *bioRxiv*, 2022–09 (2022).
15. Meyer, J. G. Qualitative and quantitative shotgun proteomics data analysis from data-dependent acquisition mass spectrometry. *Shotgun Proteomics: Methods and Protocols*, 297–308 (2021).
16. Ma, B. *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid communications in mass spectrometry* **17**, 2337–2342 (2003).

17. Cherry, J. M. *et al.* SGD: Saccharomyces genome database. *Nucleic acids research* **26**, 73–79 (1998).
18. Kim, T. K. T test as a parametric statistic. *Korean journal of anesthesiology* **68**, 540–546 (2015).
19. Fang, Z., Liu, X. & Peltz, G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023).
20. Usaj, M. *et al.* TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. *G3 Genes | Genomes | Genetics* **7**, 1539–1549. ISSN: 2160-1836. <https://doi.org/10.1534/g3.117.040220> (2022) (May 2017).
21. Van Leeuwen, J., Boone, C. & Andrews, B. J. Mapping a diversity of genetic interactions in yeast. *Current opinion in systems biology* **6**, 14–21. ISSN: 2452-3100. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6269142/> (2022) (Dec. 2017).
22. Käppeli, O. Regulation of carbon metabolism in *Saccharomyces cerevisiae* and related yeasts. *Advances in microbial physiology* **28**, 181–209 (1987).
23. Kingma, E., Diepeveen, E. T., de la Cruz, L. I. & Laan, L. Pleiotropy drives evolutionary repair of the responsiveness of polarized cell growth to environmental cues. *Frontiers in Microbiology* **14** (2023).
24. Collins, S. R., Roguev, A. & Krogan, N. J. in *Methods in enzymology* 205–231 (Elsevier, 2010).
25. Oughtred, R. *et al.* The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Science* **30**, 187–200 (2021).
26. Wilmes, G. M. *et al.* A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. *Molecular cell* **32**, 735–746 (2008).
27. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *nature* **418**, 387–391 (2002).
28. Schüller, H.-J., Richter, K., Hoffmann, B., Ebbert, R. & Schweizer, E. DNA binding site of the yeast heteromeric Ino2p/Ino4p basic helix-loop-helix transcription factor: structural requirements as defined by saturation mutagenesis. *FEBS letters* **370**, 149–152 (1995).
29. Zampar, G. G. *et al.* Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. *Molecular systems biology* **9**, 651 (2013).
30. Hammond, C., Romano, P., Roe, S. & Tontoz, P. INO2, a regulatory gene in yeast phospholipid biosynthesis, affects nuclear segregation and bud pattern formation. *Cellular & Molecular Biology Research* **39**, 561–577 (1993).
31. Hiles, I. D. *et al.* Phosphatidylinositol 3-kinase: structure and expression of the 110 kd catalytic subunit. *Cell* **70**, 419–429 (1992).
32. Weinert, T. A., Kiser, G. L. & Hartwell, L. Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. *Genes & development* **8**, 652–665 (1994).
33. Kato, R. & Ogawa, H. An essential gene, ESR1, is required for mitotic growth, DNA repair and meiotic recombination *Saccharomyces cerevisiae*. *Nucleic Acids Research* **22**, 3104–3112 (1994).

34. Kiser, G. L. & Weinert, T. A. Distinct roles of yeast MEC and RAD checkpoint genes in transcriptional induction after DNA damage and implications for function. *Molecular biology of the cell* **7**, 703–718 (1996).
35. Hector, R. E. *et al.* Mec1p associates with functionally compromised telomeres. *Chromosoma* **121**, 277–290 (2012).
36. Tkach, J. M. *et al.* Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nature cell biology* **14**, 966–976 (2012).
37. Feaver, W. J. *et al.* Genes for Tfb2, Tfb3, and Tfb4 subunits of yeast transcription/repair factor IIH: Homology to human cyclin-dependent kinase activating kinase and IIH subunits. *Journal of Biological Chemistry* **272**, 19319–19327 (1997).
38. Feaver, W. J. *et al.* Subunit interactions in yeast transcription/repair factor TFIH: Requirement for Tfb3 subunit in nucleotide excision repair. *Journal of biological chemistry* **275**, 5941–5946 (2000).
39. Marcelo, A., Koppenol, R., de Almeida, L. P., Matos, C. A. & Nóbrega, C. Stress granules, RNA-binding proteins and polyglutamine diseases: too much aggregation? *Cell Death & Disease* **12**, 592 (2021).
40. Ford, C. L., Randal-Whitis, L. & Ellis, S. R. Yeast proteins related to the p40/laminin receptor precursor are required for 20S ribosomal RNA processing and the maturation of 40S ribosomal subunits. *Cancer research* **59**, 704–710 (1999).
41. Planta, R. J. & Mager, W. H. The list of cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. *Yeast (Chichester, England)* **14**, 471–477 (1998).
42. Peng, Z.-Y., Trumbly, R. J. & Reimann, E. Purification and characterization of glycogen synthase from a glycogen-deficient strain of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **265**, 13871–13877 (1990).
43. Feng, Z. *et al.* The yeast GLC7 gene required for glycogen accumulation encodes a type 1 protein phosphatase. *Journal of Biological Chemistry* **266**, 23796–23801 (1991).
44. Cairo, L. V., Ptak, C. & Wozniak, R. W. Mitosis-specific regulation of nuclear transport by the spindle assembly checkpoint protein Mad1p. *Molecular cell* **49**, 109–120 (2013).
45. Rödel, G. Two yeast nuclear genes, CBS1 and CBS2, are required for translation of mitochondrial transcripts bearing the 5-untranslated COB leader. *Current genetics* **11**, 41–45 (1986).
46. Krause-Buchholz, U., Barth, K., Dombrowski, C. & Rödel, G. *Saccharomyces cerevisiae* translational activator Cbs2p is associated with mitochondrial ribosomes. *Current genetics* **46**, 20–28 (2004).
47. Mendenhall, M. D. & Hodge, A. E. Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews* **62**, 1191–1243 (1998).
48. Kimura, Y., Yahara, I. & Lindquist, S. Role of the protein chaperone YDJ1 in establishing Hsp90-mediated signal transduction pathways. *Science* **268**, 1362–1365 (1995).
49. Colomina, N., Gari, E., Gallego, C. & Aldea, M. The critical size is set at a single-cell level by growth rate to attain homeostasis and adaptation (2012).

50. Zarzov, P., Boucherie, H. & Mann, C. A yeast heat shock transcription factor (Hsf1) mutant is defective in both Hsc82/Hsp82 synthesis and spindle pole body duplication. *Journal of cell science* **110**, 1879–1891 (1997).
51. Brauns, F. *et al.* Redundancy and the role of protein copy numbers in the cell polarization machinery of budding yeast. *Nature Communications* **14**, 6504 (2023).
52. Wu, C.-F. *et al.* Role of competition between polarity sites in establishing a unique front. *Elife* **4**, e11611 (2015).
53. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**. Publisher: American Association for the Advancement of Science, aaf1420. <http://www.science.org/doi/10.1126/science.aaf1420> (2022) (Sept. 2016).
54. Ito, K. K., Watanabe, K. & Kitagawa, D. The emerging role of ncRNAs and RNA-binding proteins in mitotic apparatus formation. *Non-coding RNA* **6**, 13 (2020).
55. Huang, Z. *et al.* MiCroKiTS 4.0: a database of midbody, centrosome, kinetochore, telomere and spindle. *Nucleic acids research* **43**, D328–D334 (2015).
56. Rieder, C. L. Ribonucleoprotein staining of centrioles and kinetochores in newt lung cell spindles. *The Journal of cell biology* **80**, 1–9 (1979).
57. Yukawa, M., Ohishi, M., Yamada, Y. & Toda, T. The putative RNA-Binding protein Dri1 promotes the loading of kinesin-14/Klp2 to the mitotic spindle and is sequestered into heat-induced protein aggregates in fission yeast. *International Journal of Molecular Sciences* **22**, 4795 (2021).
58. Gari, E. *et al.* Whi3 binds the mRNA of the G1 cyclin CLN3 to modulate cell fate in budding yeast. *Genes & development* **15**, 2803–2808 (2001).
59. Cross, F. R. DAF1, a mutant gene affecting size control, pheromone arrest, and cell cycle kinetics of *Saccharomyces cerevisiae*. *Molecular and cellular biology* (1988).
60. Qu, Y. *et al.* Cell cycle inhibitor Whi5 records environmental information to coordinate growth and division in yeast. *Cell reports* **29**, 987–994 (2019).
61. Michel, A. H. *et al.* Functional mapping of yeast genomes by saturated transposition. *eLife* **6**, e23570. ISSN: 2050-084X. <https://elifesciences.org/articles/23570> (2021) (May 2017).
62. Shetty, A., Reim, N. I. & Winston, F. Auxin-Inducible Degron System for Depletion of Proteins in *Saccharomyces cerevisiae*. *Current protocols in molecular biology* **128**, e104 (2019).
63. Den Ridder, M., van den Brandeler, W., Altiner, M., Daran-Lapujade, P. & Pabst, M. Proteome dynamics during transition from exponential to stationary phase under aerobic and anaerobic conditions in yeast. *Molecular & Cellular Proteomics* **22** (2023).
64. Köcher, T., Pichler, P., Swart, R. & Mechtler, K. Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. *Nature protocols* **7**, 882–890 (2012).

Probing the genetic rewiring of a reproducible evolutionary trajectory

Like stones rolling down hills, fair ideas reach their objectives despite all obstacles and barriers. It may be possible to speed or hinder them, but impossible to stop them.

Jose Marti

Abstract The lack of appropriate experimental procedures in traditional wet labs to investigate the genome-wide effects of gene deletions in specific genetic backgrounds has challenged our understanding of how compensatory mechanisms work. In this study, we utilized Saturated Transposition Analysis in Yeast (SATAY) to explore the fitness effects of genome-wide mutations in the evolutionary trajectory of *bem1* mutant yeast cells, as previously reported by [1]. Our methodology allowed us to reconstruct the fitness effects of different mutations during the trajectory and examine changes in genetic interactions throughout the first step of the evolutionary process. This provided insights into how suppression mutations modify the global genetic interaction network compared to the initial deleterious mutation. Our study yields four main findings: 1) genetic interactions are highly contingent on specific experimental conditions, 2) despite experimental specificity, genetic interactors of *BEM1*, from two disparate techniques exhibit significant functional overlap, 3) suppressor mutation arrival predominantly induces sign epistasis changes in comparison to the defective mutant, and 4) spatial organization of enriched biological functions in the global genetic interaction network suggests that predicting fitness-enhancing mutations in low-fitness mutants is generally more challenging than in high-fitness mutants. These insights contribute to a better understanding of compensatory mechanisms and significantly impact personalized medicine.

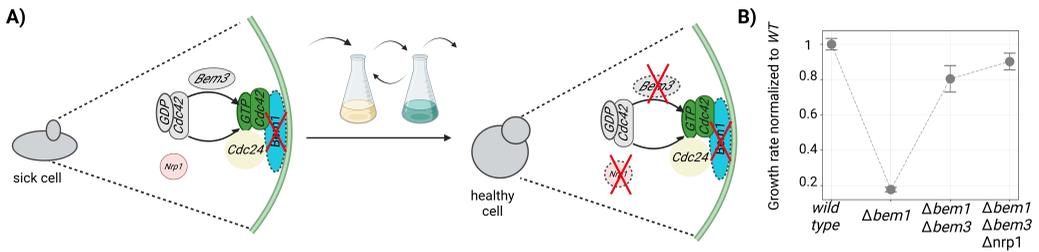


Figure 5.1. Evolutionary trajectory that recovers an impaired yeast mutant unable to establish polarization through multiple gene knockouts. **A)** A simplified model for the yeast polarity establishment consists of accumulating the complex Cdc42-Bem1-Cdc24 in one spot of the plasma membrane. This accumulation is due to a positive feedback loop centered in this complex [5, 6], that enables the directional recruitment of Cdc42-GTP to the bud site leading to bud emergence and further cell division. When Bem1 is deleted, the positive feedback loop is broken, and the capacity of the cells to polarize is highly hampered [1]. Yet, evolving *bem1* cells in rich media through a serial dilution procedure lead to the knockout of Bem3 and Nrp1 from the cell genotype. **B)** The phenotype at different steps of the evolutionary experiment, based on the population growth rate, shows a recovery of the *wild type* phenotype in the evolved cells.

5

5.1. Introduction

Cells are intricate systems of interconnected proteins that enable specific cell types to survive and function. These protein interactions have evolved to optimize the cell's required functions over millions of years. As a result, even within the same taxon, conserved functions like cell polarization may exhibit differences in their protein components [2–4]. A remarkable feature of all biological systems is their ability to adapt and reorganize their functional structure in response to external changes, such as shifts in lifestyle, food availability, and climate. This adaptability makes them evolvable systems. However, the precise mechanisms by which living systems adjust their molecular structure to accommodate external changes remain a long-standing question in biology. Understanding these mechanisms will provide insights into the flexibility and resilience of living organisms, illuminating the mysteries of evolution and adaptation.

This chapter investigates how a severely sick yeast mutant can adapt its genome structure to compensate for losing a near-essential gene. We explore how the impaired mutant evolves toward a complete functional phenotype by analyzing gene interactions. The study delves into the genetic interactions and cellular functions that characterize a crippled yeast mutant and how that changes when a loss-of-function mutation that increases fitness arises in the population.

The recovery towards a healthy mutant by losing proteins starting from a near-essential knockout gene occurred through an evolutionary trajectory, studied by [1]. The experiment involved a population of budding yeast cells with a loss-of-function mutation in the Bem1 protein, resulting in a severe fitness defect in the essential function of cell polarity. Panel A of fig. 5.1 presents a simplified depiction of the polarization network. Bem1 is a scaffold protein that initiates a positive feedback loop to recruit the Cdc24-Cdc42-GTP complex to the emerging bud site [5]. The removal of Bem1 disrupts this feedback loop, which we presume would lead to a significant deficit of Cdc42-GTP molecules reaching the bud site within the required timeframe for the cell cycle.

After 1000 generations, the population managed to recover the ability to perform cell polarization, but with three fewer proteins (Bem1, Bem3, and Nrp1), as shown in Panel B of fig. 5.1.

Interestingly, the cells regained the ability to polarize, but in a novel manner, achieved through reduced components within the cell. Another protein did not take over the original function of Bem1 in the cell [1]. Instead, there was a complete reorganization of the network. The temporal order of the mutations observed in this trajectory was reproducible, suggesting the presence of strong genetic interactions between the genes that were knocked out along the trajectory, which we presume would favor this specific trajectory over others.

To understand how these evolutionary mutations influence other genes and the overall functional reorganization in the cell, we conducted a genome-wide investigation of genetic interactions during the first two steps of the evolutionary trajectory. We utilized Saturated Transposition Analysis in Yeast (SATAY), as described in chapter 3 and chapter 4, to determine genetic interactions after *BEM1* and *BEM1BEM3* complete gene knockouts. Specifically, we computed the digenic interactions with *BEM1* and *BEM3*, as well as trigenic interactions (refer to the combined effect of three genes on a particular phenotype) with *BEM1BEM3*. We studied the main biological functions that are linked to the genes whose knockouts lead to an increase and decrease of the fitness in both genetic backgrounds. In addition, we question the SATAY analysis of genetic interactors to investigate its capability to anticipate meaningful evolutionary outcomes.

Our study yields four main findings: 1) genetic interactions are highly contingent on specific experimental conditions, 2) despite experimental specificity, genetic interactors from two disparate techniques exhibit significant functional overlap, 3) suppressor mutation arrival predominantly induces sign epistasis changes in comparison to the defective mutant, and 4) spatial organization of enriched biological functions in the global genetic interaction network (e.g. where those functions lie in the mentioned network) suggests that predicting fitness-enhancing mutations in low-fitness mutants is generally more challenging than in high-fitness mutants.

5.2. Results

5.2.1. Reconstitution of known $\Delta bem1$ suppressors using SATAY

Imperative to delineate genetic interactions for a specific species, in particular *S. cerevisiae*, is studying the fitness contribution of genetic mutations. However, high throughput techniques that enable the systematic computation of fitness effects overall coding sequences knockouts from *S. cerevisiae* are scarce. Compared to others, one technical procedure that stands out because of its technical elegance is Saturated Transposition Analysis in Yeast (SATAY). Previously, SATAY has been employed to identify novel gene functions and essentiality changes across genetic and environmental conditions. We developed a method that allows the high throughput transformation of read counts from transposon insertions along the genome to relative fitness values.

To validate that the fitness-enhancing mutations of the evolutionary trajectory were also reflected in our SATAY analysis, we generated three different SATAY mutant libraries for strains with the following mutations: *bem1-aid*, $\Delta bem3$, and $\Delta bem1\Delta bem3$. The construction of these strains is explained in section 5.4.3. Each SATAY experiment included two technical replicates, which had the same insertion profile but different read profiles due to being split before the PCR amplification step of the DNA library, as described in section 3.4. By analyzing the fitness values obtained from these libraries at each step of the evolutionary trajectory, we can verify the reliability of the SATAY data concerning the capture of the increase of fitness upon *BEM3* knockout in $\Delta bem1$ mutants and upon *NRP1* removal in $\Delta bem1\Delta bem3$ mutants.

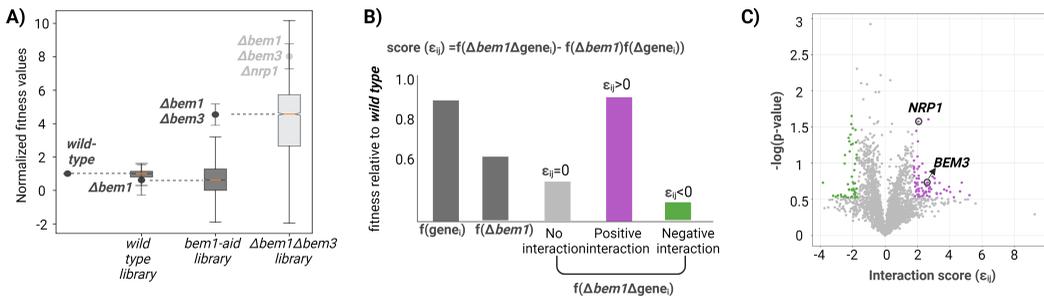


Figure 5.2. Retrieval of *BEM3* and *NRP1* as beneficial loss of function mutations of $\Delta bem1$ mutants using SATAY. **A)** Box plot indicating the distributions of fitness values from SATAY experiments in *wild type*, $\Delta bem1$ and $\Delta bem1 \Delta bem3$ genetic backgrounds. Every fitness distribution is normalized, such as its median corresponding to the fitness value of the mutants representing each background (which are symbolized with error bars). **B)** Genetic interactors are gene pairs whose combined perturbation gives rise to an unexpected phenotype, given their single effect phenotype. If the fitness of the double knockout largely increases from the product of the single mutant fitnesses, then it will indicate a positive interaction, while a noticeable reduction will hint at a negative interaction. **C)** Volcano plot of the genetic interactors of *BEM1* from SATAY experiments. In the x-axis is shown the interaction score as defined in panel A from fig. 4.5. The y-axis represents the statistical significance of the interaction based on a T-test. Purple and green dots represent significant positive and negative interactors, respectively. *BEM3* and *NRP1* are recovered as positive interactors for *BEM1*.

To reconstruct the evolutionary trajectory, we generated single, double, and triple mutant collections from the *wild type*, $\Delta bem1$, $\Delta bem3$, and $\Delta bem1 \Delta bem3$ genetic backgrounds. The single mutant collection was obtained from the same *wild type* strain described in chapter 3 and chapter 4. For the double mutant library from *bem1* mutants, we used the AID tag system, as described in chapter 4, to prevent the emergence of suppressors at early stages in the protocol. The working principle of this system is displayed in fig. 4.9. To generate the double mutant collection from *bem3* mutants, we performed a standard transformation protocol on the *wild type* strain, replacing the *BEM3* locus with the Nourseothricin antibiotic resistance cassette. The strain construction procedure details can be found in section 5.4.3. The triple mutant library was derived from the *bem1bem3* strain. Here, *BEM1* was replaced by the Geneticin (G418) antibiotic resistance cassette in *bem3::CLONAT* mutants. This approach was chosen because the fitness cost of removing *BEM1* in *bem3* mutants is not as high as in *wild type* cells. We can investigate the genetic interactions and evaluate the evolutionary trajectory by constructing these mutant collections.

We excluded the gene *BEM1* from the specific *bem1* double mutant collection during fitness computation to ensure accurate analysis. *BEM1* is a hotspot in this library because mutations in the AID tag can inactivate it, providing a full copy of *BEM1* to the population, significantly benefiting its fitness. This effect of transposon and reads enrichment in *BEM1* can be observed in fig. 5.10.

To establish a consistent baseline for fitness comparisons, we force that the fitness distribution of every mutant has a median equal to the fitness of that mutant in the background we compare it to, using a standardized procedure. Specifically, the fitness distribution of the *bem1-aid* library is centered around the fitness of *BEM1* from the wild-type library. This procedure is applied to all other mutant collections as well. The result of this process for the average fitness per gene per library is shown in Panel A) of fig. 5.2.

During the evolutionary trajectory, including the $\Delta bem1$, $\Delta bem1\Delta bem3$, and $\Delta bem1\Delta bem3\Delta nrp1$ mutants, we observed a steady increase in fitness. Each box plot in Panel A) of fig. 5.2 represents the distribution of fitness effects of a specific mutant collection, and the dots represent the individual fitness of the targeted mutants along the trajectory. The reconstitution of the suppression of *BEM3* deletion in *bem1* mutants and the suppression of *NRP1* deletion in *bem1bem3* mutants is observed.

We employed SATAY to investigate how disrupting each gene affects the organism's fitness, particularly identifying genes that can either enhance or compromise the fitness of $\Delta bem1$ mutants. Specifically, to uncover additional potential suppressors of *BEM1* besides *NRP1* and *BEM3*, we computed the digenic genetic interactors of *BEM1* and isolated the significant positive interactors of *BEM1*. To achieve this, we utilized both the *wild type* and *bem1-aid* SATAY libraries and implemented the method detailed in section 3.2.4.

For the *bem1* mutant library, we identified 5240 and 5207 suitable genes amenable to a fitness value for both technical replicates. To focus on meaningful interactors, we established criteria for selecting genes based on their interaction scores and associated p-values. Specifically, we selected genes with an absolute interaction score exceeding the mean plus one standard deviation of all interaction score values, coupled with a p-value less than 0.3. For the *bem1* mutant library, this criterion corresponded to an absolute interaction score threshold of 1.74. For the statistical significance, we set the p-value significance threshold to 0.3 due to the following reasons: 1) there is a high variability of the fitness values between the technical replicates of the *bem1-aid* library, panel A fig. 5.9, 2) the read counts exhibit a considerable bias to the gene *BEM1* as shown in fig. 5.10, 3) we lack prior information about genetic interactors of *BEM1* using SATAY, thus in this situation where little is known and the primary concern is not to miss any possible effect, a higher p-value could be temporarily acceptable to cast a wider net for detecting any effects, which would then need to be verified with more rigorous testing. It is important to notice that significance threshold of 0.3 specifically means that we are accepting a 30% of risk to have false positives in the output of this analysis, e.g. genetic interactors of *BEM1*. The subset that meet these criteria comprised 70 positive and 51 negative interactors. Such result is represented in the shape of a volcano plot (Panel C fig. 5.2). Purple and green dots represent significant positive and negative interactors based on their score value and p-value. Among the identified positive significant interactors are *BEM3* and *NRP1*, which were known suppressors from [1]. The identified interactors may shed light on potential genetic suppressors of *BEM1* and contribute to understanding its complex regulatory network.

5.2.2. Genetic interactions of *BEM1* are dependent on the specific experimental conditions

BEM1 is a gene that has been largely studied in high-throughput experiments like Synthetic Genetic Arrays (SGA), where double mutants containing a *bem1* knockout are compared to single knockouts using the population colony size as a proxy for their fitness. This method aimed to construct a global interaction network for budding yeast relying on automated yeast genetics. It tested most of the 6000 genes in the yeast *Saccharomyces cerevisiae* for all possible pairwise genetic interactions, identifying nearly 1 million interactions, including around 550,000 negative and about 350,000 positive interactions, spanning 90% of all yeast genes [7–9]. For the case of *BEM1*, this large-scale study harbors 176 and 66 significant (p-value < 0.05) negative and positive interactors of *BEM1*.

Despite the differences in technical procedures to identify genetic interactions, we explored to what extent SATAY genetic interactions predictions match the current knowledge on *BEM1* genetic interactors. This highlights how much genetic interactions are conserved or specific to the experimental conditions of the technical procedure.

Panel A),fig. 5.3, shows the difference in fitness values generated from the SGA procedure of 3206 gene deletions in the wild-type and $\Delta bem1$ genetic backgrounds. As the dotted line indicates, the normalized fitness of $\Delta bem1$ knockouts to the wild type fitness is 0.6. The significant positive and negative interactors identified during this procedure are represented by the yellow and blue dots, respectively. The immense majority of dots are in the cloud that is around the wild-type fitness (1) and the $\Delta bem1$ knockouts fitnesses, thus suggesting that genetic interactions are usually rare, as [10] pointed out. Purple and green dots highlight the significant¹ (p -value<0.3) positive and negative genetic interactors predicted by SATAY. Most of the SATAY predicted interactors fall where the neutral genes from the SGA assay are. Few negative interactors from SGA are categorized as positive interactors by SATAY, yet none of the SGA positive interactors are predicted as negative by SATAY.

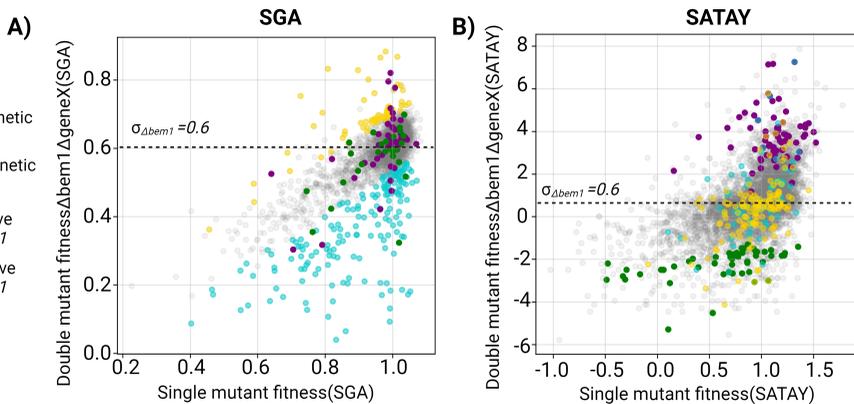


Figure 5.3. Genetic interactors are dependent on the specific experimental procedure that identifies them. **A)** Double mutant fitness in the $\Delta bem1$ background against single mutant fitness in the wild type background. Each fitness value is from the SGA assay used to construct the global genetic interaction network from [7]. Yellow and blue dots denote the significant positive and negative genetic interactors of *BEM1* from this assay. In purple and green are highlighted the significant interactors retrieved from SATAY. **B)** Double mutant fitness in the $\Delta bem1$ background against single mutant fitness in the wild type background. Each fitness value is from the SATAY assay. Purple and green dots denote the significant positive and negative genetic interactors of *BEM1* from this assay. In blue and yellow are highlighted the significant interactors retrieved from SGA. $\sigma_{\Delta bem1}$ indicates the fitness of single $\Delta bem1$ knockouts.

Panel B),fig. 5.3, depicts the differences in the predicted fitness values, from the SATAY procedure, of 5692 gene knockouts in the wild type and $\Delta bem1$ genetic backgrounds. The highest density of points is localized, as we reinforced with our normalization procedure fig. 3.8, around the fitness values of the wild type and the $\Delta bem1$ knockout predicted fitnesses. In this plot, we highlight the significant positive and negative interactors from SGA; most cluster in the neutral region. In addition, we also observed a few mismatches of the type of genetic interactors from

¹We chose a p -value < 0.3 for this specific analysis of genetic interactors of *BEM1* from the SATAY readouts for the technical replicates of the strains with the *bem1-aid* tag. We explain the choice of the specific p -value on p.135.

one method versus the other.

The fact that we found that most significant interactors predicted by SGA do not make a considerable impact on $\Delta bem1$ knockouts when using SATAY, and vice-versa, highlights how variable and diverse living cells are when subjected to different genetic and environment conditions.

Moreover, from a functional viewpoint, it is still unclear, despite the presented discrepancies, whether there is an overlap of the associated functions of the predicted interactors from both techniques.

5.2.3. Glycosylation, cell wall, vesicle traffic , signalling and transcription related modules are conserved functions of *BEM1* interactors, from SGA and SATAY procedures

The overlap between the genetic interactors from techniques like SGA and our predictions with SATAY was shown to be poor. We attributed this finding to the discrepancies between experimental conditions and yeast strain backgrounds. However, we are interested, as well, in whether the functions of the genetic interactors from both techniques are also vastly different from the functional viewpoint. If we found utterly different functional enrichment between both sets, then the associated function of *BEM1* in the network will be distinct depending on the technique used to identify it indirectly.

To implement the functional enrichment in both sets of genetic interactors from SGA and SATAY, we employed the same methodology followed by the SGA interactors displayed in the CellMap project (<https://thecellmap.org/>) by [7, 8]. The CellMap project aims to construct a functional wiring diagram for yeast cells based on a global genetic interaction similarity profile. This also provides a resource for predicting gene and pathway function of uncharacterized genes by guilt-by-association approaches. To construct the global network, they connect genes with similar genetic interaction profiles (according to a Pearson correlation measure), such that genes exhibiting more similar profiles are located closer to each other. In contrast, genes with less similar profiles are positioned farther apart in the network. Spatial analysis of functional enrichment was used to identify and color network regions enriched for similar Gene Ontology bioprocess terms. The relative position between the clusters indicates the shared functionality among the genes that compose them, see fig. 5.11; for example, DNA replication and repair (DRR) and mitosis (M) are distinct but related to chromosome segregation. Hence, in this map, they are near each other. The acronyms in the cluster names represent the following modules: S.S - MVB sorting and RIM signaling, C - Cytokinesis, C.P - Cell Polarity, Px - Peroxisome, tRNA - tRNA wobble modification, G.C - Glycosylation and cell wall, V.T - Vesicle Traffic, P.F - Protein Folding, R.B - Ribosome Biogenesis, r/nc RNA - rRNA and ncRNA processing, mRNA - mRNA processing, Ch - Chromatin, T - Transcription, N.T - Nuclear Transport, D.R.R - DNA Replication and Repair, M - Mitosis, P.T - Protein Turnover, Mtb - Metabolism, Mtc - Mitochondria. We also name the mentioned terms as the functional modules of the cell.

In the Cellmap study [7], the researchers found that negative genetic interactions often connect genes with shared functions, allowing for predictions using alternative functional data. Conversely, positive interactions, which are less functionally informative, may offer insights into mechanisms of genetic suppression or resilience.

Figure 5.4, Panel **A**), presents an overlay of the functional enrichment of *BEM1* positive interactors identified through SGA and SATAY, depicted in yellow and magenta, respectively. SGA positive interactors predominantly cluster around vesicle traffic and the interface between the

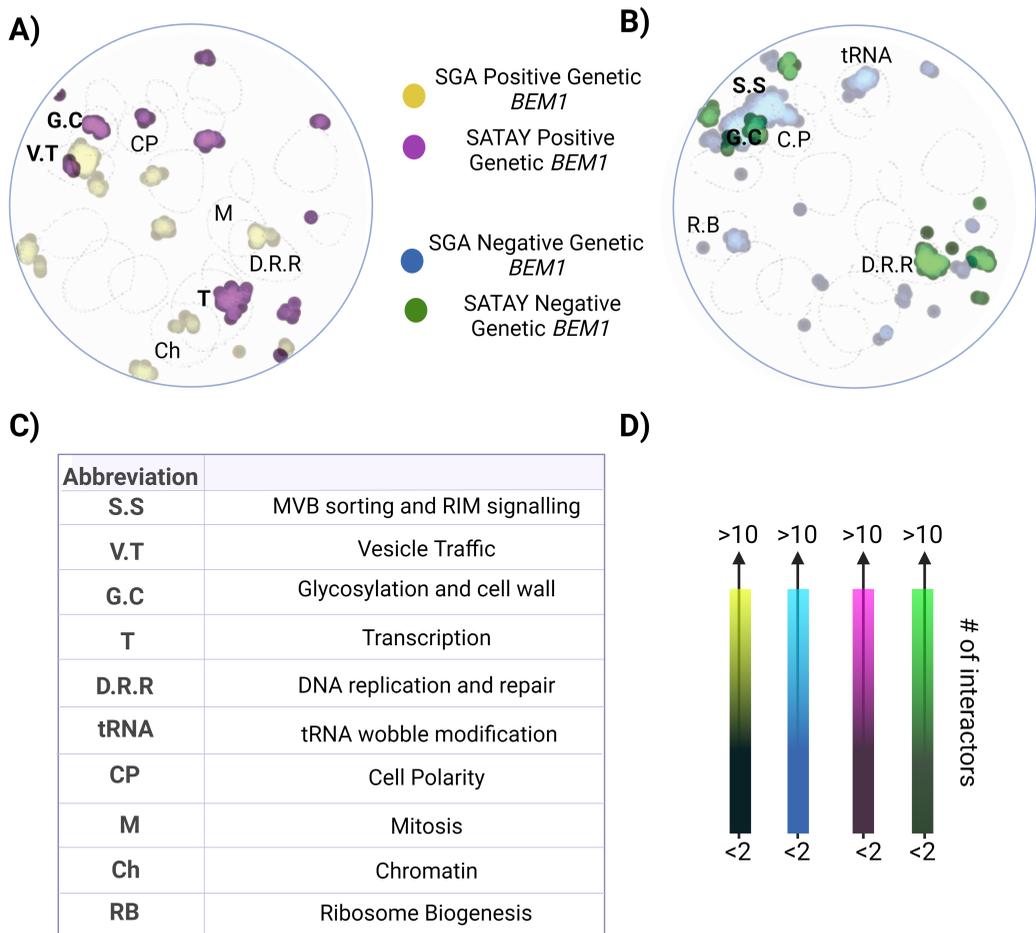


Figure 5.4. Glycosylation, cell wall, vesicle traffic, signalling and transcription related modules are conserved functions of *BEM1* interactors, from SGA and SATAY procedures. **A)** Overlay of the enrichments of the predicted significant positive genetic interactors from SGA (yellow) and SATAY (purple). **B)** Overlay of the enrichments of the predicted significant negative genetic interactors from SGA (blue) and SATAY (green). **C)** Table showing the meaning of each abbreviation used for the functional enrichment terms. **D)** Lighter colors mean a higher number of genetic interactors with a similar interaction profile. Bold functions indicate that they are common across SGA and SATAY within the same interaction type of *BEM1* interactors.

mitosis and DNA replication and repair modules, indicative of processes related to DNA segregation. Additionally, some interactors are enriched at the interface of transcription and chromatin-related modules. Surprisingly, SATAY predicted significant positive interactors also cluster around vesicle traffic and transcription-related processes. Additionally, there is a shared enrichment with the SGA counterparts in the module of glycosylation and cell wall. However, for the case of SGA interactors, they only reside in the interface with the vesicle traffic module. This observation reinforces the potential linkage of *BEM1* with vesicle traffic and transcription modules, possibly through a compensatory mechanism involving the knockout of member proteins within these functions.

In Figure 5.4, Panel **B)**, we present the overlay of functional enrichment for *BEM1* predicted sig-

nificant negative interactors from both SGA and SATAY, depicted in blue and green respectively. SGA negative interactors predominantly cluster in the signaling, glycosylation, cell polarity, tRNA, and ribosome biogenesis modules. The SATAY-predicted negative interactors of *BEM1* exhibit an overlap with SGA negative interactors in the signaling and glycosylation and cell wall-related modules. Additionally, the DNA replication and repair module shows significant enrichment, specific to SATAY. This suggests that mutations within the DNA replication and repair module may notably decrease the fitness of $\Delta bem1$ mutants, a phenomenon only discernible through SATAY experiments. Further experimental investigations are warranted to test this hypothesis.

Despite major disparities in fitness values between positive and negative genetic interactors identified through SGA and our method, resulting in variations in genetic interaction scores, our observations have unveiled overlapping outcomes regarding the functional relationship of positive and negative genetic interactions with *BEM1*. This implies that the functional linkage of *BEM1* with modules such as glycosylation, cell wall, signaling, vesicle traffic, and transcription remains consistently preserved across different strain backgrounds and growth media conditions.

5.2.4. Sign epistasis dominates the genetic rewiring in $\Delta bem1$ mutants after a *BEM3* loss of function mutation

Epistasis refers to the conditional selective effect of a mutation based on the genetic background in which it appears. While epistasis is widely observed in nature, our understanding of its consequences for evolution by natural selection remains incomplete[10]. The initial step involves studying how the total set of interactors for a specific gene is contingent upon the absence of other genetic loci. Gaining insights into how genetic interactors change in response to significant mutations, such as evolutionarily relevant mutations, could provide valuable information for exploring suppression mechanisms at the genetic interaction level. Specifically, we aim to address the impact of removing one of the well-studied suppressors of *BEM1*, such as *BEM3* [1], at the genetic interactors level. This aspect has not been studied in previous SGA studies involving trigenic mutations [11]. This serves as the overarching motivation for this section, where we explore the fitness and consequently, the genetic interaction changes at the genome-wide level of the double mutant $\Delta bem1\Delta bem3$ in comparison to the single mutant $\Delta bem1$, utilizing the capabilities of SATAY.

We followed the same strategy as [11], panel A) fig. 5.5 to compute the trigenic genetic interactions of $\Delta bem1\Delta bem3$ double mutants since it involves three genes. In the cited study, the authors measured trigenic interactions using SGA. In this procedure, they conducted 422 trigenic interaction screens, which generated around 460,000 yeast triple mutants for trigenic interaction analysis. However, this dataset does not cover interactors of *BEM1*. For the case of *BEM3*, they only performed the trigenic analysis on the double mutants $\Delta bem3\Delta ho$ and $\Delta bem3\Delta htd2$ and also did not fully explore all the genes for the third mutation. Hence, conducting a SATAY experiment to probe the gene deletion effects on the mutants of interest greatly benefits the existing data on digenic and trigenic interactions in budding yeast.

The mathematical expression for the trigenic interactions is shown in eq. (5.1). The trigenic interaction score encodes the contributions from every digenic interaction between gene X and each of the query genes (in our case: *BEM1* and *BEM3*).

$$\tau_{ijk} = f_{ijk} - f_i f_j f_k - \epsilon_{ij} f_k - \epsilon_{ik} f_j - \epsilon_{jk} f_i \quad (5.1)$$

Where f_{ijk} is the fitness of the tripled mutant, and f_i is the fitness of the $gene_i$, in the wild type background. ϵ_{ij} is the interaction score of $gene_i$ and $gene_j$.

SATAY experiments have a particularity: the mutants are generated de novo during the procedure. Thereby, the fitness values of each of the members of the constructed library are relative to the median value of the library fitness. This leads to considering the inherent nature of the directionality of genetic interactions. Directionality means that the effect of deleting gene B in $gene A$ mutant, compared to the $gene A$ mutant phenotype, may not be homologous to the opposite case because the reference phenotype has changed. For example, for the case of *BEM1* and *BEM3*, the former is not a suppressor of the latter, thus the positive interaction between them is only for *bem1* mutants, see fig. 5.14. For the mathematical expression of genetic interactions, ideally, we must compute the fitness of the double deleted as the average fitness among the two possible combinations of double deleted mutants, namely $geneA geneB$ and $geneB geneA$. Generally, we only have data for one combination of the double mutants; for instance, to compute the fitness of *bem1 geneX*, we would require SATAY data also in *geneX* background, which we typically don't have. The only case where this is fulfilled for our dataset is for computing the fitness of *bem1bem3* mutants. For simplicity, we will assume that $f_{ij} = f_{ji}$ for the cases we don't have enough data. To compute the trigenic interactions to *BEM1* and *BEM3*, we are limited to the constructed SATAY datasets in the backgrounds: *WT, bem1-aid, Δbem3, Δbem1 Δbem3*. Thereby, we can only base our calculation on the combinations: 1) $i \rightarrow$ gene of interest, $j \rightarrow BEM1$, $k \rightarrow BEM3$, 2) $i \rightarrow$ gene of interest, $j \rightarrow BEM3$, $k \rightarrow BEM1$.

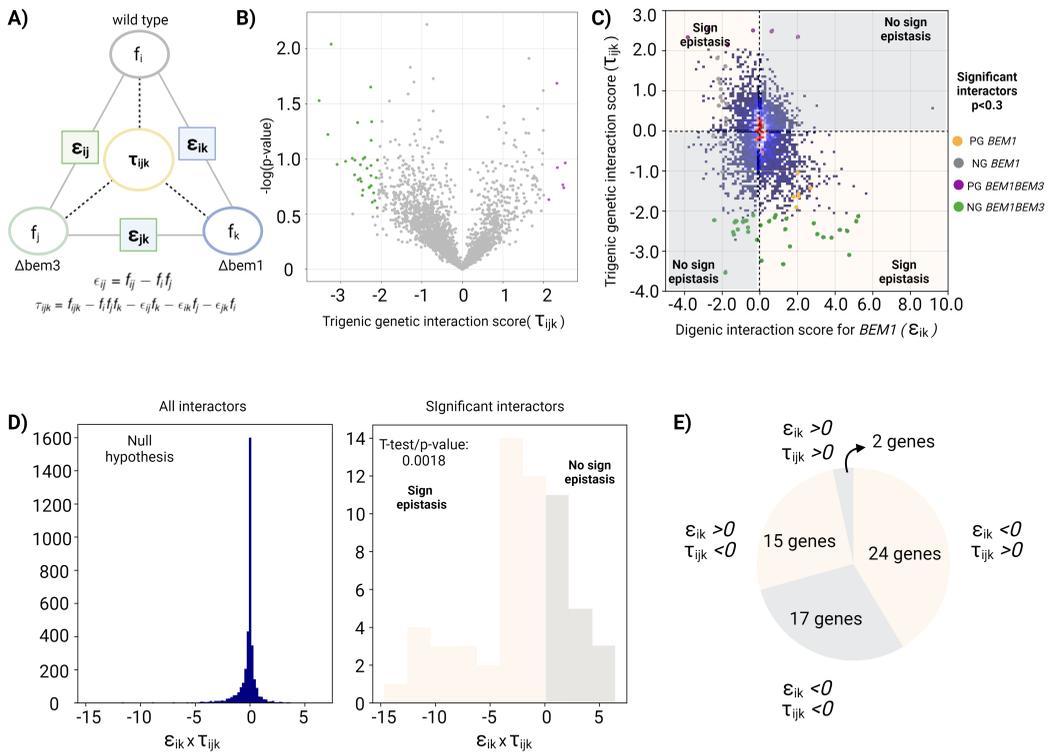


Figure 5.5. Sign epistasis dominates the changes of interaction from $\Delta bem1$ to $\Delta bem1\Delta bem3$ backgrounds. **A)** Trigenic genetic interactions are influenced by all the digenic interactions between their gene pairs. The computation of the trigenic interactors in $bem1bem3$ cells requires the calculation of the digenic interactors with $bem1$ and $bem3$. f refers to the fitness values of the population in different genetic backgrounds. ϵ refers to the digenic genetic interaction score, namely, between two genes. τ refers to the trigenic genetic interaction score, namely, between three genes. **B)** Volcano plot of $BEM1BEM3$ predicted trigenic interactions, the x-axis displays the value of the score and the y-axis the $-\log(p\text{-value})$. Green and purple dots show the predicted negative and positive significant interactors. **C)** 2D Histogram of the genetic interaction scores of the predicted interactors of $BEM1BEM3$ (y-axis) against the predicted interactors of $BEM1$ (x-axis). Yellow and gray squares denote the area where sign epistasis occur or not, according the sign of the genetic interaction scores from each background. Significant interactors of $BEM1$ are highlighted in yellow and gray, for positive and negative, respectively. Significant positive and negative trigenic interactors of $BEM1BEM3$ are depicted in magenta and green respectively. **D)** The distribution of the product of digenic and trigenic interaction scores from significant interactors (right) is significantly different ($p\text{-value}=0.001$) than the same distribution with all interactors (left). ϵ_{ik} refers to the digenic genetic interaction score with $BEM1$ and τ_{ijk} refers to the trigenic genetic interaction score with $BEM1BEM3$. **E)** Pie plot that quantifies the changes of the significant genetic interactors types from $BEM1$ to $BEM1BEM3$.

Those combinations lead to the computation of an average trigenic interaction score, eq. (5.2).

$$\tau_{ijk} = \frac{1}{2}(\tau_{i10} + \tau_{i01}) \quad (5.2)$$

$$\tau_{\text{gene},\text{bem1bem3}} = \tau_{i01} = f_{i01} - f_i f_0 f_1 - \epsilon_{i0} f_1 - \epsilon_{i1} f_0 - \epsilon_{01} f_i \quad (5.3)$$

$$\tau_{\text{gene},\text{bem3bem1}} = \tau_{i10} = f_{i10} - f_i f_1 f_0 - \epsilon_{i1} f_0 - \epsilon_{i0} f_1 - \epsilon_{10} f_i \quad (5.4)$$

Putting eq. (5.3) and eq. (5.4) in eq. (5.2), and assuming that $f_{i01} = f_{i10}$ we get that :

$$\tau_{ijk} = f_{i01} - f_i f_0 f_1 - \epsilon_{i0} f_i - \epsilon_{i1} f_0 - \frac{1}{2} f_i (\epsilon_{01} + \epsilon_{10}) \quad (5.5)$$

Computation of the digenic interaction scores between geneX and *BEM1* or *BEM3* involves the computation of the fitness of geneX deletion in the *bem1-aid* and Δbem3 SATAY libraries. In the fig. 5.12(from Appendix) is depicted a 2D histogram comparing the extracted fitness distributions between the Δbem1 and the $\Delta\text{bem1}\Delta\text{bem3}$ genetic backgrounds. Dotted lines denote the predicted fitness values of the corresponding mutants, namely, Δbem1 from the wild type library and $\Delta\text{bem1}\Delta\text{bem3}$ from the Δbem1 library. The red color in the histogram shows the denser region of dots, indicating the median normalization we performed on the fitness dataset to ensure that most gene knockouts do not impact a certain phenotype and thereby reinforce the notion that genetic interactions are rare. This is an assumption that we extend from the wild-type background. However, in different genetic backgrounds, it is still unclear if this remains the case.

To compute the trigenic genetic interaction score, we used the technical replicates of the $\Delta\text{bem1}\Delta\text{bem3}$ mutants to average over and calculate their statistical significance. Panel B),fig. 5.5, shows the volcano plot representing the outcome of the genetic interaction calculation. We obtained 3563 genes amenable to compute their trigenic interaction scores based on the intersection of suitable genes for their fitness computation between the wild type, *bem1-aid* and Δbem3 SATAY libraries. We defined significant genetic interactions as genes with an average score greater/lower than the $\text{mean} + \text{std} / -(\text{mean} + \text{std})$ of all scores distribution, and its p-value is at most 0.3. This method renders 6 and 32 significant positive and negative genetic interactors of *BEM1BEM3* respectively. Intuitively, this can be explained by analyzing the shape of the fitness distribution of mutants in the $\Delta\text{bem1}\Delta\text{bem3}$ background. The fitness distribution is left skewed to fitness values below the computed fitness value of $\Delta\text{bem1}\Delta\text{bem3}$ mutants, fig. 5.12, thus the chances to find significant positive interactors, e.g. suppressors are remarkably low. The low number of significant negative interactors is mainly attributed to the cumulative noise between the technical replicate libraries of all genetic backgrounds implicated in this procedure.

Next, we explored how the genetic interactions in both genetic backgrounds were compared, panel C) fig. 5.5. This 2D histogram consistently shows that most genetic interactions cluster around zero in both backgrounds. However, many genes that appear as potential suppressors for *BEM1* have a negative interaction score when *BEM3* is knocked out. The opposite also holds; there is a considerable number of potential negative interactors of *BEM1* that become positive when *BEM3* gets removed. This phenomenon is called sign epistasis. The significant genetic interactors from both environments are highlighted in different colors, shown in the legend.

To assess the significance of the sign epistasis phenomenon on the significant genetic interactors, we compare the distribution of the product of digenic and trigenic genetic interaction scores for the significant interactors with the distribution for all interactors. The distribution for all interactors is centered on zero (fig. 5.12D), with tails for positive and negative values. We conduct a T-test to determine if the distribution of significant interactors differs significantly from

that of all interactors, which serves as the null hypothesis. Our analysis reveals a p-value of 0.001, indicating a significant difference between the two distributions. Thus, we conclude that the sign epistasis phenomenon is indeed significant for the set of genetic interactions under evaluation.

To visualize the quantification of the changes of gene interaction types from the $\Delta bem1$ background to the $\Delta bem1\Delta bem3$ background, see fig. 5.5 E). In the $\Delta bem1\Delta bem3$ genetic background, 39 significant genes alter their interaction score, constituting approximately 67% of the total 58 significant genes.

We argue that this observation helps elucidate the significant phenotypic shift observed in $\Delta bem1\Delta bem3$ mutants compared to $\Delta bem1$ mutants. However, it's essential to acknowledge the limited statistical power of these experiments.

However, what are the enriched functional modules of the significant genetic interactors for *BEM1BEM3*? And how do the antagonistic and conserved genetic interactor types from *BEM1* to *BEM1BEM3* translate to the functional modules from the global genetic interaction network? The following section will address these questions.

5.2.5. Glycosylation, cell wall and transcription related functions are epistatic to *BEM3* in $\Delta bem1$ mutants

It holds significant implications for gene therapy within the realms of cancer research and personalized medicine to understand the impact of specific mutations on different genetic backgrounds present in various tumors [12, 13]. This knowledge is crucial for identifying mutations that might eliminate a particular tumor but, if the tumor harbors a different mutation, could promote its proliferation. More importantly, comprehending not only the genes but also the biological functions, where alterations affect tumor viability contingent on the specific genetic background, is essential for designing more precise and targeted therapies for patients.

In the preceding section, we show that sign epistasis characterize the change of genetic interactions from $\Delta bem1$ to $\Delta bem1\Delta bem3$ background. One example of such a gene is *FLO11*, whose knockout in $\Delta bem1$ mutants may lead to cell death whereas the same mutation, if *BEM3* is additionally removed, might be beneficial for the cell's fitness. *FLO11* is involved in flocculation, coflocculation, cell adhesion during biofilm formation, pseudohyphal growth, and invasive growth; it localizes to the plasma membrane, the bud neck, and the extracellular region [14]. This predicts that the ability to perform flocculation and facilitate cell adhesion during biofilm formation may be essential for $\Delta bem1$ mutants, while it is not critical for $\Delta bem1\Delta bem3$ double mutants.

To have a general overview of the enrichment differences in antagonistic genetic interaction types, we used the same method as for fig. 5.4, which is taken from the Cellmap project [7]. Its general principle is to cluster genes that have a similar genetic interaction profile. The relative position between the functional clusters indicates the shared functionality among the genes that compose them. fig. 5.6 shows the enrichment for all the possible combinations of genetic interaction types of *BEM1* and *BEM1BEM3*.

In Panel A of fig. 5.6, we illustrate the functions enriched for predicted significant negative *BEM1* genetic interactions and their antagonistic interaction type for *BEM1BEM3*. A notable co-enrichment is observed in glycosylation and cell wall-related processes. Alterations in these functions may have both detrimental or beneficial effects depending on the presence or absence of *BEM3* in $\Delta bem1$ mutants, suggesting functional epistasis in these genetic backgrounds. Interestingly, *FLO11* is a cell-surface glycoprotein, but its precise location remains uncertain due to its limited genetic interaction similarity with other network proteins.

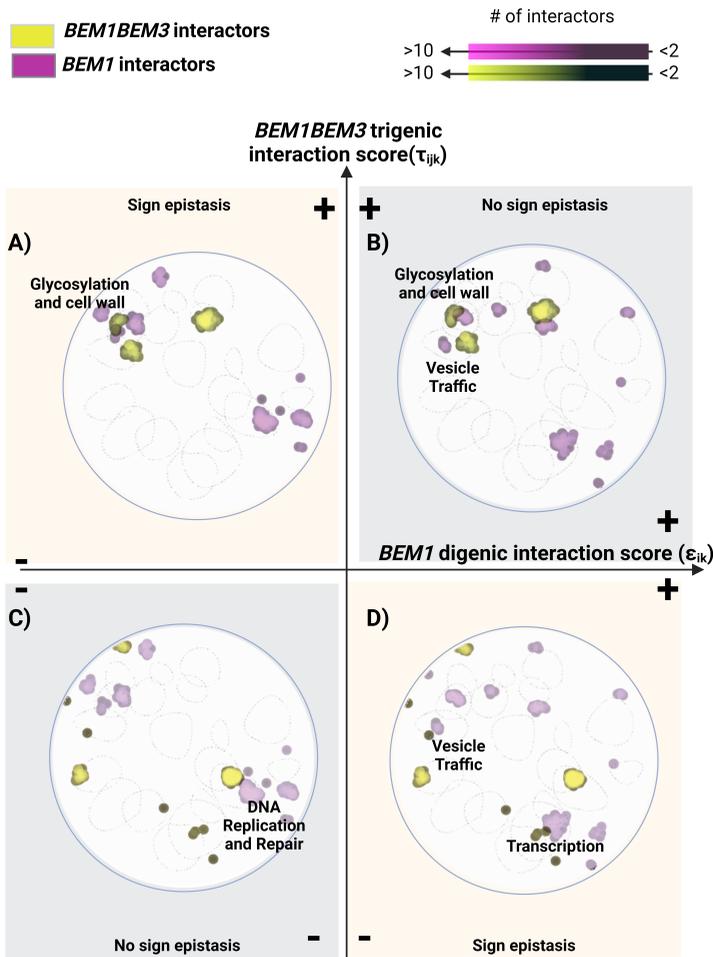


Figure 5.6. Glycosylation, cell wall, vesicle traffic and transcription related functions are epistatic to *BEM3* in $\Delta b e m 1$ mutants. Yellow and magenta dots symbolize functional enrichment of *BEM1BEM3* and *BEM1* interactors, respectively. The color bar indicates the number of interactors with similar genetic interaction profiles with our target genes, with lighter colors indicating a higher number. **A)** X-axis reflects the predicted significant *BEM1* negative genetic interactors and y-axis shows the predicted significant *BEM1BEM3* positive genetic interactors. **B)** X-axis reflects the predicted significant *BEM1* positive genetic interactors and y-axis shows the predicted significant *BEM1BEM3* positive genetic interactors. **C)** X-axis reflects the predicted significant *BEM1* negative genetic interactors and y-axis shows the predicted significant *BEM1BEM3* negative genetic interactors. **D)** X-axis reflects the predicted significant *BEM1* positive genetic interactors and y-axis shows the predicted significant *BEM1BEM3* negative genetic interactors.

Remarkably, the enrichment analysis for *BEM1* positive interactors, as depicted in Panel D of fig. 5.6, reveals the presence of the glycosylation and cell wall module. This finding is consistent with the observation from SGA predicted positive *BEM1* interactors shown in Panel A of fig. 5.4. It underscores the dynamic nature of functional modules, which are not static regions with fixed protein sets. Instead, proteins within a module can serve multiple shared functions through their genetic or physical interactions with other cellular proteins. For instance, the glycosylation and cell wall module involves a post-translational modification that plays various structural and

functional roles in membrane and secreted proteins [15]. Additionally, it can influence the thermodynamic and kinetic stability of proteins [16].

Conversely, in our investigation of antagonistic interactions with *BEM1BEM3* negative interactions, we discovered that mutations affecting vesicle traffic and transcription-related functions could confer benefits to $\Delta bem1$ mutants while simultaneously compromising fitness if *BEM3* is knocked out.

Regarding the functional modules that maintain their influence on the fitness of both $\Delta bem1$ and $\Delta bem1\Delta bem3$ double mutants, vesicle traffic, glycosylation, and cell wall-related processes appear to be consistent targets of compensatory mutations in both scenarios. Notably, vesicle traffic was weakly enriched for negative *BEM1BEM3* interactors as well. This observation is intriguing considering that vesicle traffic plays a crucial role in cell polarization, which is closely related to these mutants. Polarized growth involves the transport, tethering, and fusion of secretory vesicles with the plasma membrane at the bud site, facilitated by transport along actin cables oriented toward the bud site [17]. Therefore, mutations enhancing this pathway may provide an alternative mechanism for polarization when the reaction-diffusion pathway is compromised due to the deletion of core proteins.

Additionally, mutations affecting the DNA replication and repair module exhibit deleterious fitness effects in both genetic backgrounds, as shown in Panel C) fig. 5.6.

In summary, our findings indicate that mutations impacting the glycosylation, cell wall, vesicle traffic, and transcription functional modules demonstrate sensitivity to the presence or absence of *BEM3* in $\Delta bem1$ mutants. Notably, alterations within various components of the glycosylation and cell wall modules can either support the survival or lead to the demise of $\Delta bem1$ mutants, regardless of whether *BEM3* is present. Moreover, the transcription module emerges as a potential rescuer of $\Delta bem1$ mutants, albeit at the potential cost of reduced fitness if *BEM3* is also absent.

5.2.6. The spatial distribution of the functional enrichment of genetic interactors renders contrasting differences between $\Delta bem1$ mutants and $\Delta bem1\Delta bem3$ mutants

Throughout this chapter, we have seen one example of how different functions are reshaped based on the genetic makeup of the cell. It is fair to mention that the mutations we are considering are part of the same "module" that is the bud emergence module (BEM), however they display very distinct interaction and phenotypic pattern in the yeast cell.

Biological processes are necessary interconnected to maintain the optimal physiological state of a cell, and these connections can be influenced by various factors like the environment and genetic alterations. Often, in biological studies aimed at characterizing specific proteins using different mutants [18], the broader impact of perturbing one element on the functions of others is overlooked. This can lead to a misleading perception of orthogonality among proteins with diverse roles in the cell, supporting the idea of a modular design for cellular functions[19].

In this modular paradigm, functions are considered as sets of interacting components that operate independently of one another, each with its own dynamics. However, this paradigm typically does not explicitly address how mutations in one set can alter the dynamics of other sets.

Throughout this chapter, we've explored an example of how different functions can be reshaped based on the genetic composition of the cell. Based on our findings, we suggest an

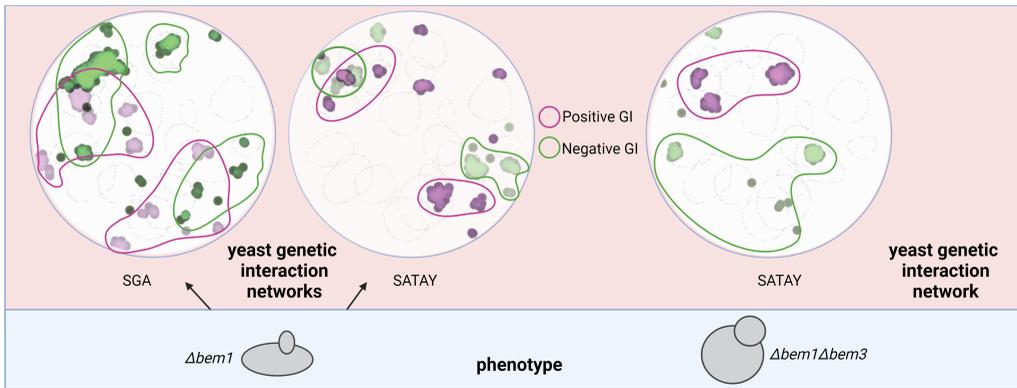


Figure 5.7. Spatially different functional enrichment of $\Delta bem1$ (Left) and $\Delta bem1\Delta bem3$ (Right) interactors. Functional enrichment of the positive and negative genetic interactions are encircled in magenta and green colors, respectively.

5

alternative approach to observe the transition in functions following a nearly essential mutation and subsequent compensatory mutation. This approach involves examining the overlap and physical proximity on the yeast interaction map of the functionally enriched sets associated with both types of genetic interactions. Our hypothesis posits that a larger degree of overlap and/or proximity within the functions associated with both types of interactors suggests a fuzzy fitness effect of altering those functions to the specific mutant. In such a network, the traditional notion of modularity may not adequately explain how each function affects the overall fitness of the cell.

In fig. 5.7, we present a visual illustration of this perspective, unfolding from the phenotype to the functional enrichment of the predicted yeast genetic interaction networks. On the left, it illustrates the scenario of the $\Delta bem1$ crippled phenotype. By employing SATAY (or SGA), we can construct the genetic interaction network, categorizing positive and negative genetic interactors of the gene knockout in question. Analyzing the overlap in the functional enrichment of these genetic interactors from each category allows us to speculate on the challenge of predicting which functional module to modify to enhance the fitness of the specific mutant.

We observed from the functional enrichment analysis of genetic interactors for *BEM1* and *BEM1BEM3*, that there is an overlap or close proximity between the functions associated with negative and positive genetic interactors of *BEM1*. Specifically, we found that glycosylation, cell wall, and vesicle traffic functions are shared between both types of interactions, see panel A and B from fig. 5.6. Additionally, we observed positive interactors enriched in the transcription module and negative ones in the DNA replication and repair module, which are closely located in the map, suggesting a potential sharing of components or pathways between them. This result emphasizes the intertwined nature of altering these functions concerning the cell's fitness, for this particular mutant. As a result, we suggest that predicting strong compensatory mutations for these mutants from our study might be inherently challenging due to this intricate interplay.

In contrast, genetic interactors of *BEM1BEM3* appear to be functionally isolated from one another, with considerable physical separation on the map. As a result, mutations that would enhance the fitness of the double mutant, harboring a compensatory mutation for *BEM1*, are distinctly different from those that would impair its fitness. It's worth noting that we assume

the yeast genetic network of higher-order interactions follows a similar organizational pattern as the digenic yeast interaction network. Thus, any enrichment resulting from computing similar genetic profiles is solely due to digenic interactions. Considering this, we propose that the fitness landscape may be significantly smoother and, therefore, more predictable than that of the compromised mutant. Furthermore, we noted that the functional enrichment of the genetic interaction network involving the gene *BEM3* lacks a noticeable overlap across different types of genetic interactions, as illustrated in fig. 5.13. $\Delta bem3$ mutants display a phenotype similar to the wild type one, see fig. 5.14 and Panel A and B from fig. 5.15.

In summary, our findings support the notion that the modular perspective of the cell is valuable for understanding its normal physiological state. However, when the cell encounters external stress, whether genetic or environmental, the applicability of this framework may become less evident. In the next section, we will explore the accuracy of predicting evolutionarily significant mutations using our method of identifying significant genetic interactions from SATAY data.

5.2.7. SATAY predicted compensatory mutations in the $\Delta bem1$ genetic background confirmed results from a studied evolutionary trajectory

Based on our previous sections, we propose that SATAY-derived predictions of genetic interactions can identify genes and functions that may act as potential compensatory or lethal mutations, thereby contributing to the pool of mutations upon which natural selection can act during evolution. In the next section, we will explore a well-documented experimental evolutionary trajectory involving $\Delta bem1$ mutants, as detailed in [1], to assess the efficacy of SATAY in predicting mutations along this trajectory.

The genetic makeup of an evolving population, representing a specific evolutionary outcome, is influenced by various factors such as the dynamic environment (the evolving population can also change the environment) and the fitness landscape, which is represented by the intricate relationships between genotype-phenotype and phenotype-fitness maps [20–22]. The inherent complexity of these properties in evolving populations renders the task of predicting long-term evolution challenging and nearly unattainable [23].

Specially, we would like to refer to forward-looking evolutionary predictions at the genotypic level. For example, predicting which genes will acquire new mutations or confer increased fitness to a microbial population carrying a harmful mutation. In this scenario, environmental and genetic factors will influence the predictability of evolutionary outcomes, namely, unpredictable environmental changes, epistasis, and population size [23].

Despite the myriad factors influencing evolutionary processes, we aim to assess the predictive capacity of genetic interactors identified by SATAY as potential mutational targets for a specific evolutionary trajectory. This prediction relies solely on the fitness advantages conferred by knocking out certain genes in a yeast population. We will leverage the mutational trajectory measured by [1], which involves the recovery of fitness in a yeast population with a full *BEM1* knockout. During this trajectory, fifteen additional mutations, including *BEM3*, *NRP1*, and *BEM2*, were detected at various time points but were not fixed in the population, panel A) fig. 5.8. One hypothesis is that these mutations were not fixed because they did not provide a significant fitness advantage to sweep through the population. Regarding the hypothetical order of non-fixed mutations, one plausible scenario is that they occurred after the deletion of *BEM3*. This speculation arises from the fact that the *BEM3* stop codon mutation appeared early in the experiment,

around 30 generations of yeast division cycles. Consequently, we anticipate observing a compensatory effect of these mutations in the *bem1bem3* mutant in our data. An essential assumption in our analysis is the treatment of all mutation types as full knockout mutations, as we posit that transposon insertions within a gene predominantly result in gene disruptions.

Upon examining the genetic interaction scores of genes harboring mutations along the evolutionary trajectory, we identified 7 mutations out of the possible 12^2 that exhibit potential suppression properties in the $\Delta bem1$ background, panel B) fig. 5.8. Five of the seven predicted SATAY suppressors of *BEM1* are statistically significant. Despite that the environmental conditions from the evolution experiment and the SATAY procedure are substantially distinct, in terms of media and incubation conditions, we obtained that more than half of the mutations along the trajectory can also rescue $\Delta bem1$ mutants. Interestingly, some of them show a stronger suppression power than *BEM3* deletion.

However, none of those mutations seems to increase the fitness of $\Delta bem1\Delta bem3$ mutants. Most of the mutations show a deleterious effect in this background. We attributed this result to the limitation of SATAY to recover small fitness increases of an already fit mutant like $\Delta bem1\Delta bem3$. Thus, we speculate that maybe the suppressive power of these mutants to $\Delta bem1\Delta bem3$ is not sufficiently high to be captured by SATAY experiments. We may also think that most of these mutations arose after *NRP1* deletion thus, they are epistatic to *NRP1*, which means being beneficial when *NRP1* is removed and the opposite otherwise. However, this is not applicable for the case of the four genes *GPR1*, *TCB1*, *IRA1* and *GAL83* since on the lines they occurred, the *NRP1* mutation was not present.

On the other hand, as shown in previous section, it is possible that specific genes do not match our predictions due to environment and genetic differences between the experiments, yet, the functional modules to which those genes belong may coincide, as shown for the *BEM1* interactors from SGA and SATAY techniques. Thus, we explored whether the functional enrichment of the mutations along the evolutionary trajectory aligns to the functional enrichment of predicted positive interactors for *BEM1* and *BEM1BEM3*.

The mutations identified along the evolutionary trajectory cluster in processes primarily related to ribosomal (r)RNA/non-coding (NC) RNA processing, transfer (t)RNA wobble modification and to cell signaling and polarity, panel C fig. 5.8. There is a slight functional co-enrichment observed with the significant positive interactors of *BEM1*, as identified in the SATAY analysis. This co-enrichment is situated close to transfer (t)RNA wobble modification and polarity related functions. Remarkably, these functions did not have an overlap with the functional enrichment of negative genetic interactors of *BEM1*, shown in panels B and C of fig. 5.6. In addition, the functional enrichment for positive interactors of *BEM1BEM3* does not exhibit any co-localization with the enrichment of mutations along the trajectory.

It is insightful that the primary functions behind the mutations that rescue $\Delta bem1$ mutants, from the evolutionary trajectory, are related to the protein synthesis machinery at various levels (ribosomal (r)RNA/non-coding (NC) RNA processing and transfer (t)RNA wobble modification) rather than being directly localized to cell polarity, for example. This suggests that suppression can act more globally, influencing all functions in the cell, rather than specifically targeting the function to which *BEM1* is assigned to.

To summarize, through the SATAY genetic interaction analysis of *BEM1* and *BEM1BEM3*, we

²There were three genes: *ERV29*, *SOI4*, *YKL091C* that were not amenable for the SATAY fitness calculation, due to that their transposon reads are less than five

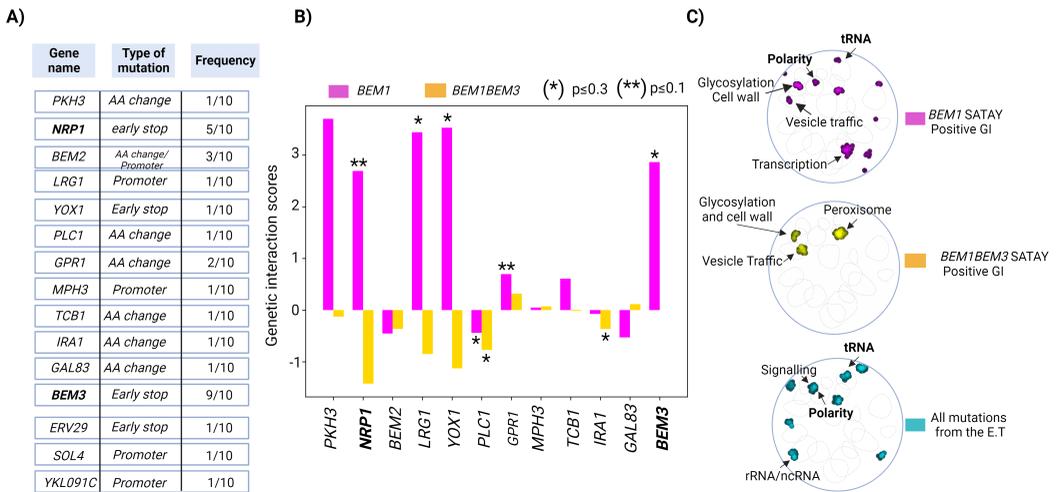


Figure 5.8. SATAY predictions based on genetic interactions only recover potential *BEM1* suppressors. **A)** Mutations that show up along the evolutionary trajectory (E.T) from [1] and their type. **B)** Bar plot depicting in the y-axis the genetic interaction score value and in the x-axis the recovered mutations along the evolutionary trajectory. The interaction scores with *BEM1* are shown in magenta and with *BEM1BEM3* in yellow. The asterisk on top of the column signifies that that specific genetic interaction value has a p-value below 0.3. **C)** Functional enrichment over the global genetic interaction map of *BEM1*, *BEM1BEM3* predicted significant positive Genetic Interactors (GI) and mutations over the evolutionary trajectory.

could delineate the significant compensatory effect of the non-fixed mutations along the evolutionary trajectory that recovers $\Delta bem1$ mutants from [1]. Despite the proposed higher predictability in the double mutants to identify compensatory mutations, shown in fig. 5.7, it is highlighted that our SATAY analysis was unable to accurately identify the compensatory effects of these mutations in the $\Delta bem1 \Delta bem3$ background. We attributed this result to the presumable weak suppression effect of these mutations and the inability of our specific SATAY assays to distinguish those as significant interactors in the double mutant background.

5.3. Summary and discussion

Throughout this chapter, we studied the genetic rewiring that occurs in an evolutionary trajectory that recovers a crippled yeast mutant by inactivating two extra genes in 1000 generations [1]. In particular, the sick mutant is due to the knockout of the gene *BEM1*, and the evolved mutant contains *BEM3* and *NRP1* inactivated by a stop codon mutations, respectively. We pursued to study the genetic rewiring along the first step of this trajectory because it is when the population's major phenotype changes occur, see fig. 5.1.

Before the study of the genetic rewiring embodied in the major abrupt phenotypic change along the trajectory, we shed light on the tight dependence of the genetic interactions with the environment and genetic backgrounds. We illustrated that genetic interactions identified by a different technique, SGA, are predominantly neutral genes, in a setup like SATAY. SGA and SATAY high throughput techniques largely differ on their experimental procedure, concerning media types, incubation times, and strain backgrounds. All these factors were also thought to greatly impact the outcome of essential genes determination [24]. In our study, we confirm that genetic

interactions are also subjected to these crucial properties. However, we exhibited that despite the mentioned differences, both sets of positive and negative interactions manifest conserved functional enrichment. We show that glycosylation, cell wall, vesicle traffic and transcription related modules were common enrichments along the predicted positive interactors from both techniques. Glycosylation, cell wall and signalling related modules were shared as well by the predicted negative interactors from both techniques, panels B and C fig. 5.4. Remarkably, both methods agreed on the simultaneous role of certain functions like glycosylation and vesicle traffic on benefiting or harming the fitness of $\Delta bem1$ mutants.

The enrichment analysis conducted in the global study of suppressors in yeast [24] revealed that processes related to chromosome segregation, endocytosis, glycosylation, cell wall, protein folding, and protein degradation are prominently involved in compensating for severe growth defects associated with impairment of the polarity and morphogenesis module. This study extensively examined literature-curated suppression interactions, identifying approximately 6000 potential interactions from 1700 published papers sourced from the BioGRID's "synthetic rescue" dataset. Notably, our study also found associations with most of these functions among the significant predicted positive interactors of *BEM1* despite the variety of technical procedures used to investigate genetic interactors from the different papers and mutations affecting the polarity and morphogenesis module.

On the other hand, the comparison of functional enrichment in predicted genetic interactors of *BEM3* from both techniques does not reveal consistent characteristics as observed for *BEM1*, as shown in fig. 5.13 in the Appendix. This discrepancy prompts inquiries into the relationship between the fitness of a gene knockout and its predictive genetic interaction network. The findings suggest that if a knockout mutation has a neutral or minimal detrimental effect on fitness, its genetic interaction network may exhibit greater variability across experimental procedures. Conversely, mutations significantly impairing cell fitness tend to have more robust and consistent interactions with their genetic partners. The relationship between fitness effects and genetic interaction network variability remains intriguing and warrants further investigation through additional experiments.

However, the identification of specific genetic interactors is strongly dependent on the chosen p-value cutoff. We selected a moderate cutoff of 0.3 based on three primary considerations: significant variability among technical replicates, the absence of prior information about SATAY-predicted *BEM1* and *BEM1BEM3* genetic interactors, and a notable bias in read-counts towards *BEM1* in the *bem1-aid* mutant library, which resulted in many transposons across the genome receiving zero or few reads. In particular, adopting a smaller p-value cutoff, such as 0.1, would lead to identifying only 20 genes for further analysis in the case of the *bem1-aid* mutant library, and no genes for the double mutant library. Therefore, we needed to find an optimal balance between the quantity and statistical quality of the predicted genetic interactors. For future experiments, it is highly advisable to increase the number of replicates to reduce the impact of variability due to random factors and outliers.

Furthermore, the knowledge of possible gene and functions that may compensate or aggravate fitness contribute to the understanding of the genetic rewiring that occur during an evolutionary trajectory. Our study reveals that most genes (approximately 80%) exhibit neutrality in response to Bem1 and Bem3 inactivation within the genome. However, it's important to note that this result is subjected to limitations imposed by the statistical power of our SATAY datasets. When examining the predicted genetic interactors in the $\Delta bem1 \Delta bem3$ background, we observe

that around 67% (39 out of 58 genes) of them change their interaction type when transitioning from $\Delta b e m 1$ to $\Delta b e m 1 \Delta b e m 3$ (sign epistasis), as shown in panel E of fig. 5.5. We attribute this observation to the significant genetic rewiring that accompanies a drastic phenotypic change, such as the one occurring when comparing $\Delta b e m 1$ to $\Delta b e m 1 \Delta b e m 3$ mutants. This insight represents a step forward in exploring the impact of suppressor mutations on the broader interaction network concerning genes affected by the initial deleterious mutation. It raises the question of whether the cell, in its efforts to recover from the initial harmful mutation, must reconfigure a significant portion of its genetic interactions.

Furthermore, we demonstrated that glycosylation, cell wall, vesicle traffic, and transcription represent epistatic functions for both the *BEM1* and *BEM3* genes. Notably, glycosylation and cell wall-related functions may exhibit both beneficial and detrimental effects on the fitness of $\Delta b e m 1$ mutants, regardless of the presence of *BEM3*, see fig. 5.6A-B. This suggests that proteins associated with glycosylation and cell wall may engage in diverse physical and genetic interactions, resulting in multifaceted effects for this particular genetic background.

In the case of the $\Delta b e m 1 \Delta b e m 3$ background, we note that certain vesicle traffic-related gene knockouts exhibit improvement in fitness (fig. 5.6B), while a subset of few genes from the same module may detrimentally affect the fitness of such mutants (fig. 5.6C).

An intriguing finding from our study is the functional segregation among genetic interactors of *BEM1BEM3*. Notably, the enrichment patterns of antagonistic interaction types are distinctly separated on the map, as shown in the right map from fig. 5.7. This suggests that positive and negative interactors may operate within separate pathways, with some overlap observed within the vesicle traffic module. This stands in contrast to our observations regarding *BEM1* interactors, where we observed functional overlap between opposite interaction types. Even when there was no direct overlap, as seen in the case of transcription for positive interactors, it was still in proximity on the map to DNA replication and repair-related functions of negative interactions.

The divergence observed in the functional relationships of genetic interactors may provide insights into the characteristics of the fitness landscape underlying these genetic backgrounds. In cases where mutants carry a nearly essential mutation, predicting compensatory mutations based on the landscape can be an extremely challenging task. Conversely, for recovered mutants, the predictability of compensatory mutations could be considerably higher. This finding suggests that it is easier to differentiate the effects of mutations in the normal physiological state of a yeast cell compared to a compromised state where a shift in functionalities across different modules appears to be necessary for its survival. This highlights the complex interplay of genetic interactions and functional modules in response to genetic perturbations, especially in scenarios where cells are under stress due to genetic knockouts.

Conversely, our method showed higher accuracy in predicting evolutionary-relevant mutations compensating for $\Delta b e m 1$ mutants than for $\Delta b e m 1 \Delta b e m 3$ mutants, as previously identified in [1], as depicted in Panel A of fig. 5.8. However, SATAY also predicted some mutations to have a greater compensatory effect than *BEM3* itself, possibly due to environmental variations across experiments, as shown in Panel B of fig. 5.8. Notably, there was no overlap between the predicted positive interactors of *BEM1BEM3* and the functionally enriched mutations from the trajectory, as illustrated in Panel C of fig. 5.8. This outcome may be attributed to the specific limitations of our SATAY assay in identifying significant positive interactions for an already relatively fit mutant.

Considering all the findings collectively, we propose that the genetic rewiring observed after a compensatory mutation, which restores the fitness of a sick mutant, can be characterized by

a transition towards a more 'modular' scenario. In this state, individual functions have a distinct and predictable impact on the cell's fitness, resembling the behavior seen in the cell's normal physiological state when a specific function is altered. Essentially, the fitness landscape that defines the condition of a sick mutant shifts from being non-deterministic and rugged to becoming more deterministic and smoother. One intriguing question that remains is whether all compensatory mutations lead to the same effect in sick mutants, or if this observed pattern is an exception rather than a rule.

5.4. Materials and Methods

5.4.1. Materials

List of strains

Name	Genotype	Source
yLIC133a	MATx can1-100,leu2-3,112 ,his3-11,15, ura3Δ,ade2Δ BUD4-S288C	This study
yLIC137a	MATx can1-100,leu2-3,112 ,his3-11,15, ura3Δ,ade2Δ, bem3::CLONAT ,BUD4-S288C	This study
yTW001	MATx can1-100,leu2-3,112 ,his3-11,15, ura3Δ,ade2Δ bem3::CLONAT ,bem1::KanMx,BUD4-S288C	Tomas Wisse Master Thesis (our lab)
yWT003	MATx can1-100leu2-3,112 ,his3-11,15 , ura3Δ,ade2Δ , ho::osTIR1-3xMyc-kanMX, bem1::mCherry-AID,BUD4-S288C	Wessel Theunisse Master Thesis (our lab)
yWT004	MATx can1-100leu2-3,112 ,his3-11,15 , ura3Δ,ade2Δ , ho::osTIR1-3xMyc-kanMX, bem1::mCherry-AID, Δbem3::CLONAT, BUD4-S288C	Wessel Theunisse Master Thesis(our lab)
yLL117	MATx can1-100,leu2-3,112 ,his3-11,15, ura3Δ, bem3::CLONAT ,BUD4-S288C	Laan Lab

Table 5.1. List of strains used in this chapter.

List of primers

Name	Sequence
oLIC51	GCATATATAATGTTGTCTTGAACCCAC
oLIC52	GTCTGACTAGGTAAAGTACCAGATAAGAACCG
oLIC54	GAGAACGGCATCACATCTGGGG
oLIC55	GCGCCACAATGAAATCTACGCCAC
oLIC56	GCGGATAAGTCAAGCATCCATTG
oLIC57	GGGACTCACATCTATCTTGGGCC
oLL29	AAAAGGACAATTACAAACAGGAATCGA
oLL30	AATCACGAATGAATAACGGTTTGGT
oLL401	GACATATTTGTCGAGGGAGGTGAT
oLL402	CTATCGTCATAAGTTGGTGTATTTGCG

Table 5.2. List of primers used in this chapter.

5.4.2. SATAY libraries

Strain	Nickname	# of transposons	# of reads	# of transposons in CDS	# of reads in CDS
yLIC137_a	Δ bem3	598686	54731167	333492	4040247
yLIC137_b	Δ bem3	645482	58018144	357018	6067071
yTW001_a	Δ bem1 Δ bem3	496985	65988828	254192	2730809
yTW001_b	Δ bem1 Δ bem3	471995	64317775	233686	2590647
yWT003_a	bem1-aid	78731	14798259	48437	96192
yWT003_b	bem1-aid	138538	15471164	82167	187579

Table 5.3. Characteristics of the SATAY libraries used in this chapter.

Media

Standard yeast culturing. Standard culturing was performed in YPD (10g/L Yeast extract, 20g/L Peptone, 20g/L dextrose), SC (6.9 g/L Yeast nitrogen base, 0.75 g/L Complete supplement mixture, 20g/L dextrose). All liquid media were filtered and sterilized to prevent degradation of the media components, especially amino acids.

Non-standard culturing media. For ade^- strains, the growth media was supplemented with 22mg adenine per 400 mL batch of media during the SATAY procedure. For culturing the *bem1-aid* strains, 0.1M stock solutions of indole-3-acetic acid (IAA) were prepared by dissolving 175mg

IAA in 2mL of 100% absolute ethanol. The final volume was adjusted to 10ml, adding extra ethanol when necessary to prevent precipitation of IAA. The stock solution was filter sterilized using a 0.2 μm syringe filter, aliquoted, and stored at -20°C .

5.4.3. Methods

Strain construction. yLIC137 and yTW001 were constructed using the homologous recombination strategy. At first, yLIC133 was transformed, using a high-efficiency protocol ([25]), with an amplified DNA fragment from a PCR with yWT004 gDNA and primers oLIC51 and oLIC52. Details of the transformation can be found in <https://leilaicruz.github.io/Experimental-journal-jupyterB00K/journal/2020-11/2020-11-27-gDNA-extraction-and-PCR-with-ywt04a.html> and <https://leilaicruz.github.io/Experimental-journal-jupyterB00K/journal/2021-01/2021-01-25-Checking-dbem3-dbem3dnrp.html>. The resulting strain was yLIC137, which was checked by sequencing using primers oLL401 and oLL402. The same strategy was applied to construct yTW001. In this case, yLL117 gDNA was extracted, and the *BEM1::KanMX* cassette was amplified with the primers oLIC54 and oLIC55. The resultant DNA was transformed, using the same procedure described in [25] into yLIC137 to yield yTW001. Precise sequencing was performed to examine the absence of *BEM1* and *BEM3* loci using the primers oLIC56, oLIC57, oLL29, and oLL30.

Spatial enrichment analysis with thecellmap.org. To reproduce panel A), B) from fig. 5.4, and fig. 5.6, please download the list of genes in <https://github.com/SATAY-LL/comparative-analysis-among-tree/main/publishing>, and select the set of interactors you want to visualize. Then go to the web application tool (thecellmap.org.) and press the button named "Overlay data" then go to the "Upload a list of genes" and paste the list of genes there. Adjust the color in the lower bar at the bottom left and click Submit.

5.5. Appendix

Appendix: Fitness data noise between technical replicates of *bem1-aid* and $\Delta bem1\Delta bem3$ libraries

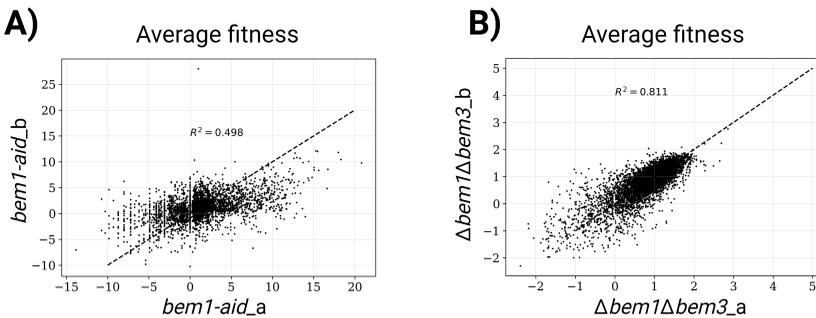


Figure 5.9. Noise between the fitness of each gene from individual technical replicates. **A)** Average fitness of the technical replicates from the *bem1-aid* strain. The R^2 signifies the person correlation among the replicates. **B)** Average fitness of the technical replicates from the $\Delta bem1\Delta bem3$ strain. The R^2 signifies the person correlation among the replicates.

Appendix: Overrepresentation of *BEM1* in the SATAY data, using the AID-tag system

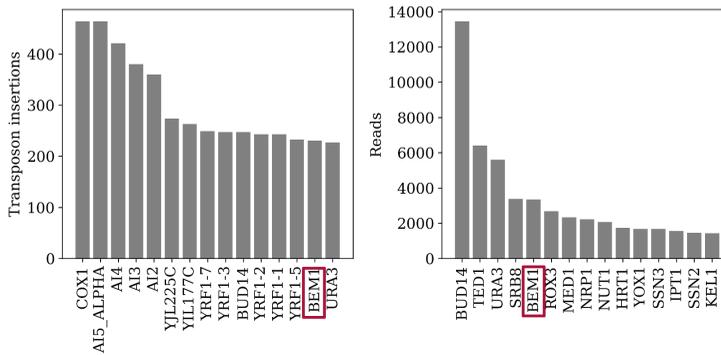


Figure 5.10. Top 15 more enriched genes with transposon and reads from the strain containing the AID-tag system attached to *BEM1*.

5

Appendix: Cellmap enrichments

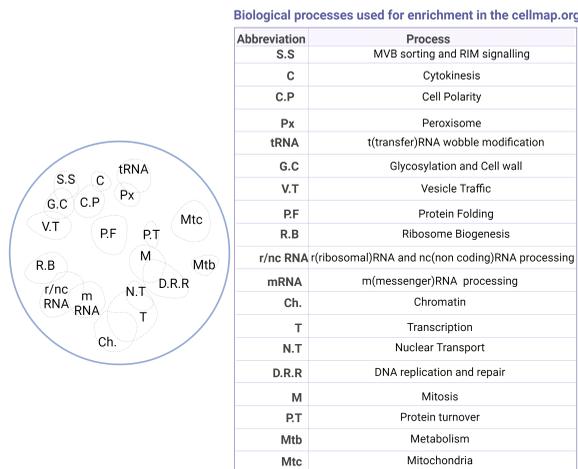
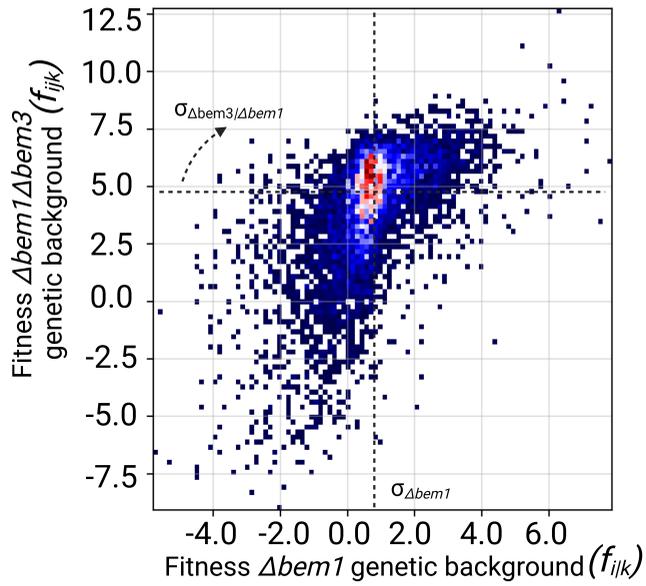


Figure 5.11. Functional clustering of budding yeast biological processes adapted from [8].

Fitness differences across *BEM1* and *BEM1BEM3* predicted SATAY genetic interactors



5

Figure 5.12. 2D Histogram of the fitness values in the $\Delta bem1 \Delta bem3$ (y-axis) against the $\Delta bem1$ genetic backgrounds. $\sigma_{\Delta bem1}$ and $\sigma_{\Delta bem1 \Delta bem3}$ are the SATAY predicted fitness values of those mutants.

BEM3 predicted fitness landscape based on SGA and SATAY

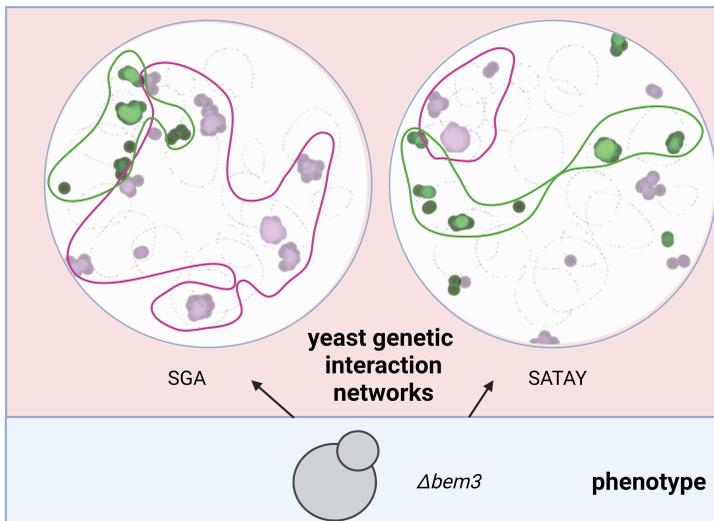


Figure 5.13. Predicted fitness landscapes of $\Delta bem3$ mutants. Functional enrichment of the positive and negative genetic interactions are encircled in magenta and green colors, respectively. The "F" letters indicate the functions corresponding to each cluster. The 2D representation of the fitness landscape is shown, where the x-axis represents alterations of different functional modules, and the y-axis represents the fitness corresponding to such alterations.

Appendix: Directional nature of genetic interactions with SATAY

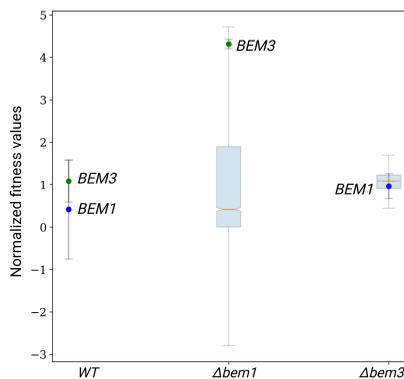


Figure 5.14. Directional character of genetic interactions with SATAY. Case of *BEM1* and *BEM3* interactions. The x-axis represents three different genetic backgrounds for SATAY libraries, and the y-axis is the normalized fitness values per library regarding the median of the library and the fitness value of the target mutant. It is shown that *BEM1* deletion is not a suppressor for *bem3* mutants, while the opposite holds.

Fitness landscapes and genetic interactions changes of $\Delta bem3$ mutants compared to $\Delta bem1\Delta bem3$

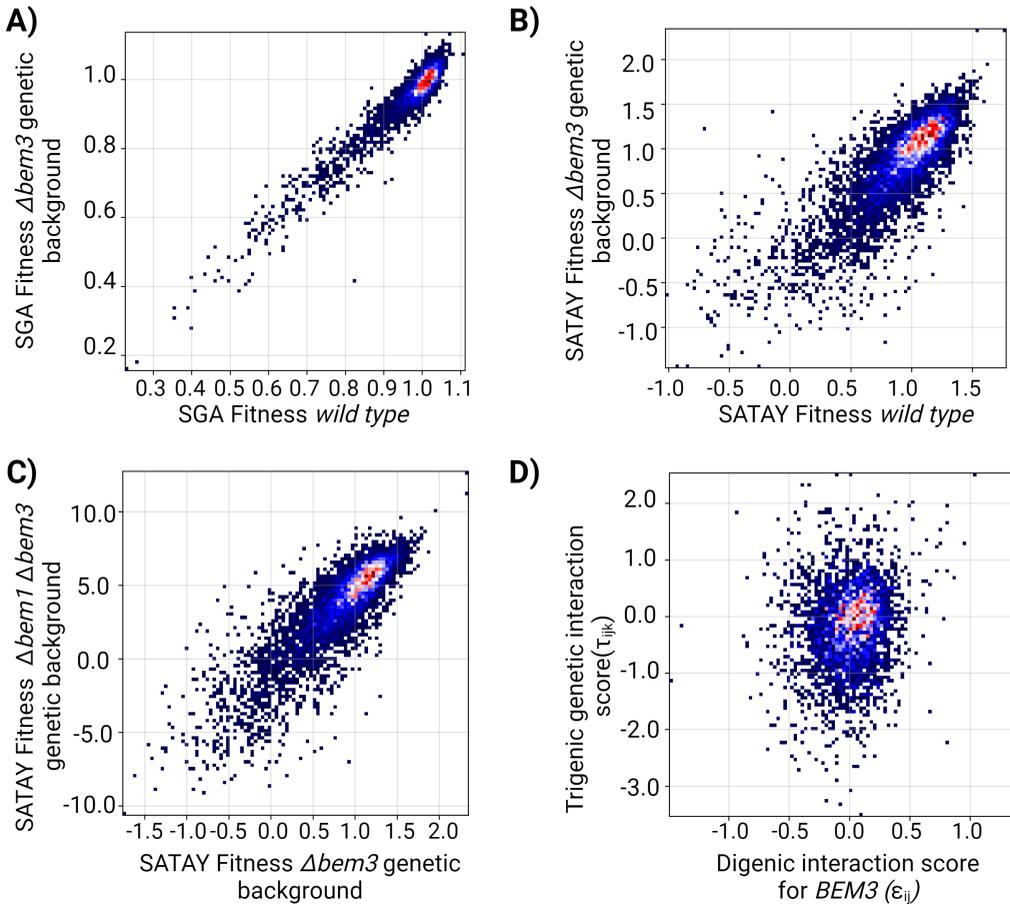


Figure 5.15. Consistency of fitness changes after $BEM3$ deletion in wild type from SATAY and SGA experiments, and genetic interaction changes after deleting $BEM1$ in $BEM3$ mutants. **A)** 2D Histogram of the fitness values in $\Delta bem3$ and wild type genetic backgrounds from SGA measurements. **B)** 2D Histogram of the fitness values in $\Delta bem3$ and wild type genetic backgrounds from SATAY data. **C)** 2D Histogram of SATAY fitness values in $\Delta bem3 \Delta bem1$ and $\Delta bem3$ genetic backgrounds. **D)** 2D Histogram of SATAY predicted genetic interactors for $BEM1BEM3$ and $BEM3$.

Bibliography

1. Laan, L., Koschwanez, J. H. & Murray, A. W. Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *Elife* **4**, e09638 (2015).
2. Diepeveen, E. T., de la Cruz, L. I. & Laan, L. Evolutionary dynamics in the fungal polarization network, a mechanistic perspective. *Biophysical Reviews* **9**, 375–387 (2017).
3. Diepeveen, E. T., Gehrman, T., Pourquié, V., Abeel, T. & Laan, L. Patterns of conservation and diversification in the fungal polarization network. *Genome Biology and Evolution* **10**, 1765–1782 (2018).
4. Bergmiller, T., Ackermann, M. & Silander, O. K. Patterns of evolutionary conservation of essential genes correlate with their compensability. *PLoS genetics* **8**, e1002803 (2012).
5. Klünder, B., Freisinger, T., Wedlich-Söldner, R. & Frey, E. GDI-mediated cell polarization in yeast provides precise spatial and temporal control of Cdc42 signaling. *PLoS computational biology* **9**, e1003396 (2013).
6. Brauns, F. *et al.* Adaptability and evolution of the cell polarization machinery in budding yeast. *bioRxiv*, 2020-09 (2020).
7. Usaj, M. *et al.* TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. *G3 Genes | Genomes | Genetics* **7**, 1539–1549. ISSN: 2160-1836. <https://doi.org/10.1534/g3.117.040220> (2022) (May 2017).
8. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**. Publisher: American Association for the Advancement of Science, aaf1420. <http://www.science.org/doi/10.1126/science.aaf1420> (2022) (Sept. 2016).
9. Costanzo, M. *et al.* The genetic landscape of a cell. *science* **327**, 425–431 (2010).
10. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).
11. Kuzmin, E. *et al.* τ -SGA: synthetic genetic array analysis for systematically screening and quantifying trigenic interactions in yeast. *Nature protocols* **16**, 1219–1250 (2021).
12. Roth, J. A. & Cristiano, R. J. Gene therapy for cancer: what have we done and where are we going? *Journal of the National Cancer Institute* **89**, 21–39 (1997).
13. Cross, D. & Burmester, J. K. Gene therapy for cancer treatment: past, present and future. *Clinical medicine & research* **4**, 218–227 (2006).
14. Cherry, J. M. *et al.* SGD: Saccharomyces genome database. *Nucleic acids research* **26**, 73–79 (1998).
15. Varki, A. *et al.* Essentials of Glycobiology [internet] (2015).
16. Ardejani, M. S., Noodleman, L., Powers, E. T. & Kelly, J. W. Stereoelectronic effects in stabilizing protein–N-glycan interactions revealed by experiment and machine learning. *Nature chemistry* **13**, 480–487 (2021).

17. Harris, K. P. & Tepass, U. Cdc42 and vesicle trafficking in polarized cells. *Traffic* **11**, 1272–1279 (2010).
18. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nature methods* **11**, 801–807 (2014).
19. Melo, D., Porto, A., Cheverud, J. M. & Marroig, G. Modularity: genes, development, and evolution. *Annual review of ecology, evolution, and systematics* **47**, 463–486 (2016).
20. De Visser, J. A. G. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**, 480–490 (2014).
21. Fragata, I., Blanckaert, A., Louro, M. A. D., Liberles, D. A. & Bank, C. Evolution in the light of fitness landscape theory. *Trends in ecology & evolution* **34**, 69–82 (2019).
22. Wright, S. *et al.* The roles of mutation, inbreeding, crossbreeding, and selection in evolution (1932).
23. Wortel, M. T. *et al.* Towards evolutionary predictions: Current promises and challenges. *Evolutionary applications* **16**, 3–21 (2023).
24. Van Leeuwen, J. *et al.* Exploring genetic suppression interactions on a global scale. *Science* **354**. Publisher: American Association for the Advancement of Science, aag0839. <http://www.science.org/doi/10.1126/science.aag0839> (2022) (Nov. 2016).
25. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature protocols* **2**, 31–34 (2007).

Conclusions

Biological systems exhibit complex and dynamic behavior. This complexity stems from the interplay of multiple levels of organization (gene, protein, functions, phenotype, population), making it challenging to understand their functioning comprehensively. Moreover, their dynamic nature is characterized by the continuous evolution and adaptation of internal components in response to the ever-changing external environment.

Despite the intricate complexity inherent in living organisms and the apparent inability to predict evolutionary outcomes [1–3], certain experiments demonstrate the reproducibility of evolutionary trajectories [4, 5]. These experiments emphasize the significance of epistatic interactions, which can constrain various evolutionary pathways [6, 7]. Epistasis¹ is recognized as a potent concept with the potential to forecast evolutionary outcomes[8–10].

In this work, we study the implications of epistatic interactions on how biological functions emerge from the collective interplay of numerous components. In particular, we focus on the cell polarization machinery of *Saccharomyces cerevisiae* as a model system. In budding yeast, this function has been shown to depict a rapid and reproducible recovery of polarization ability following the knockout of a key component, Bem1. Through mathematical modeling and experimental validation, we unveil latent mechanisms within the cell-polarization machinery that shed light on the redundancy and adaptability of cellular polarization networks.

Moreover, we leverage the potential of a high-throughput methodology known as Saturated Transposition Analysis in Yeast (SATAY) to investigate the impact of these epistatic interactions on the genetic rewiring of the cell. This approach elucidates how a suppressor mutation impacts the overall cellular machinery. We aim to decipher the underlying mechanisms governing compensatory evolution by integrating diverse methodologies to probe the effects of suppressors at both functional and genome-wide scales.

¹Epistasis refers to a phenomenon in genetics where the effect of one gene (or genetic variant) is modified by the presence or absence of alleles at another gene or genetic locus. In other words, it describes the interaction between different genes that influence the expression of a particular trait. Epistasis can manifest in various ways, including masking the effects of one gene by another, amplifying or diminishing the effects of a gene, or creating novel phenotypic outcomes that are not predictable based on the individual effects of each gene. Understanding epistasis is crucial for unraveling the complexities of genetic inheritance patterns and phenotypic diversity within populations.

Redundant self-organization mechanisms coexist and are responsible for the resilience and adaptability of the cell polarity machinery.

The adaptability of biological functions is exemplified in the Cdc42 cell-polarization machinery of budding yeast, where cells swiftly regain their polarizing and dividing capabilities following the loss of a crucial component. Despite this remarkable phenomenon, the mechanisms underpinning this recovery have remained enigmatic.

To tackle this challenge, we undertook a comprehensive approach integrating mathematical modeling, experimental validations, and conceptual theory to delve into the intricacies of the cell polarization module in budding yeast. Our investigation honed in on the Cdc42-polarization system, unearthing three distinct self-organization patterns within the network. These patterns transcend the conventional wild-type mechanism, which hinges on the co-localization of Cdc42 with its GEF via Bem1. Instead, we unveiled a latent rescue mechanism independent of Bem1 and a mechanism operating autonomously of Cdc42 redistribution. Our theoretical framework, corroborated with experimental data, reveals that these mechanisms share numerous components and interaction pathways within the network. This suggests that redundancy in cell polarization does not merely reside at the level of individual components or interactions but rather emerges at the functional process level. Even if one submodule fails, the collective action of the remaining submodules is responsible for an operational mechanism for cell polarization. This resilience is attained by meticulously fine-tuning parameters, particularly protein copy numbers, to a regime where these remaining submodules can compensate.

From a genetic perspective, genes encoding components of the cell polarization machinery exhibit dosage sensitivity [11]. Mutations affecting cis-regulatory elements, such as promoters and enhancers, can precisely adjust protein copy numbers within a specific range crucial for the functioning of a particular cell-polarization mechanism, thus optimizing its function. However, this sensitivity to protein copy numbers also imposes constraints on the evolution of the polarization machinery's components through processes like duplication and sub-functionalization [11, 12]. Such evolutionary mechanisms may be limited by the need to maintain the delicate balance of protein dosage required for proper functioning of the cell polarization machinery.

Furthermore, our findings underscore the pivotal role of redundancy in augmenting the system's evolvability, facilitating the emergence of new functions while preserving the original ones. Scaffold proteins, notably Bem1, play a pivotal role in this process, fostering the evolution of functional modules. We posit an evolutionary trajectory for Bem1, suggesting its origins from a more rudimentary ancestral mechanism that gradually acquired the capacity to mediate current polarization mechanisms. This evolutionary path of Bem1 likely constitutes a fundamental aspect of fungal evolution, evidenced by its conservation across fungi [13] and the similarity of polarization mechanisms in related species [14, 15].

In general, our study has demonstrated how understanding the mechanistic principles underlying self-organization can offer valuable insights into the evolution of cellular functions, a fundamental aspect of evolutionary cell biology. Specifically, we have provided a concrete example illustrating how a self-organizing system may have evolved from a rudimentary, parameter-sensitive mechanism to a more specific, robust, and tightly controlled mechanism over time.

Constructing genetic interactions maps based on transposon-based mutagenesis screens.

Fitness landscapes are fundamental to predictive models of adaptive evolution, providing insights into the accessibility and likelihood of evolutionary trajectories [16–18]. Empirical fitness landscapes, constructed by systematically assessing allele combinations, have been instrumental in explaining observed evolutionary patterns. However, their application in predictive models is complicated by their dynamic nature across environments and genetic backgrounds.

For instance, studies have shown that the structure of empirical fitness landscapes can vary due to evolving cells' effects on their environment [7], and the lethality of gene loss can differ significantly between genetic backgrounds. This underscores the need for techniques that enable the construction of fitness maps in diverse environmental and genetic contexts, a feature lacking in many existing methods [19–23].

Transposon insertion sequencing (TIS) methods offer a rapid means of constructing libraries containing single gene deletion mutants within specific genetic backgrounds or environments. Initially used to identify essential genes, TIS methods have evolved to construct genome-wide fitness maps of gene disruption mutants based on mutant abundance and observed read counts.

In our study, we describe a method to estimate the fitness of gene deletion mutants using data generated from a Saturated Transposition Analysis in Yeast (SATAY) screen for *Saccharomyces cerevisiae*. Our approach, based on average read count per transposon insertion site, yields robust fitness estimates across replicate SATAY experiments (with the exception of special libraries like the *bem1-aid*). However, despite this robustness, the distribution of fitness effects (i.e. the relationship between a mutant fitness and the reference fitness) differs significantly from those generated by other studies, specifically from Synthetic Genetic Arrays (SGA) and barcode sequencing fitness estimates. This highlights the inconsistencies between fitness distributions generated with different high-throughput techniques. This phenomenon was already pointed out by [24, 25], who demonstrated that the lethality of gene loss in *S. cerevisiae* can vary substantially between can vary across species, strains, and environments. The observed conditional essentiality was attributed to a complex set of background-specific modifiers that influence the phenotype of mutations and contribute to differences in essentiality between individuals. However, one consistent feature across various techniques is that most gene deletions have neutral effects on the population [26].

In *chapter 3*, we propose a method utilizing fitness estimates from SATAY to compute statistically significant genetic interactions. Despite the variability in fitness values among technical and biological replicates in our samples, we argue that this method enables the exploration of a broader mutant space, leading to the discovery of significant and novel genetic interactions among specific genes of interest.

A challenge in using SATAY for determining fitness estimates of gene disruptions is the lack of a genuine reference for fitness values, particularly when studying different genetic backgrounds or environmental conditions. The fitness values derived from SATAY are based on the assumption that most transposon insertions within the central 80% of the gene have neutral effects on the cell. While this assumption may be valid for the wild-type genetic background, it may not hold for other genetic backgrounds or diverse environmental conditions. Addressing this challenge requires the identification of more suitable references that allow for proper fitness comparisons across different conditions. This could involve using alternative experimental methods to measure gene fitness under specific conditions and using these results as a reference to validate and calibrate the SATAY fitness values.

In conclusion, leveraging SATAY across different genetic backgrounds in budding yeast offers valuable insights into the fitness implications of disrupting various genes and detecting gene essentiality under specific conditions. However, to ensure comprehensive and accurate results, it is crucial to supplement SATAY findings with more targeted assays, such as microscopy or population growth studies of mutant yeast strains, performed under identical laboratory conditions as SATAY. These additional experiments serve to validate and enhance the fitness values provided by SATAY consistently, thereby strengthening the overall reliability and applicability of the obtained results.

Proteins with an RNA binding domain may suppress deleterious gene disruptions in cell polarity.

Proteins containing RNA binding domains (RBDs) play vital roles in diverse cellular processes by interacting with RNA molecules, influencing functions like RNA splicing, transport, stability, and translation. Among these proteins, Nrp1 is known to harbor an RBD. Nrp1 has been associated with cellular polarization in *Saccharomyces cerevisiae* but without a specific biological function. *NRP1* has been shown to have epistatic interactions with *BEM1*, a key player for the cellular polarization machinery of budding yeast. Cells harboring a knockout on *BEM1* can quickly recover after *BEM3* and *NRP1* gene deletions. This outcome is not expected from the individual phenotype that $\Delta nrp1$ mutants display.

Chapter 4 attempts to explore the possible *NRP1* biological role in budding yeast by studying the influence of *NRP1* on gene expression and the overall genetic interaction map. We presented experimental evidence indicating that Nrp1's RNA binding domain has minimal impact on gene expression at the cellular level. Our mass spectrometry analysis, covering around 1,200 proteins, revealed no significant changes in expression levels, leading us to conclude that Nrp1 does not directly regulate gene expression in budding yeast.

From SATAY experiments and the functional enrichment of the predicted genetic interaction network of *NRP1*, we envision that *NRP1* is involved in crucial cellular processes related to the cellular division machinery, such as determining cell size during the START phase of the cell cycle, exit from the G0 phase and spindle morphology. Moreover, we also see that respiratory growth is a function that *NRP1* may be involved, as observed by [27].

Recent research has underscored the involvement of RBD proteins in mitotic spindle formation across different eukaryotes. Studies suggest that disruptions to mitotic spindle structure caused by transcription inhibitors or RNase occur independently of active translation. This implies that functional mRNAs or non-coding RNAs (ncRNAs) might act as regulators or structural components of the mitotic spindle. In fission yeast, the ortholog of *NRP1*, known as *DRI1*, has been directly linked to mitotic spindle assembly. Notably, the absence of *DRI1* results in reduced kinesin-14/Klp2 protein levels on the spindle, suggesting a potential rescue mechanism for $\Delta cut7$ mutants. Furthermore, *DRI1*'s involvement in heterochromatin assembly via its RNA binding domain is noteworthy. Our analysis did not reveal significant genetic interactions between *NRP1* and the *CUT7* orthologs *KIP1* or *CIN8* in budding yeast. However, a minor, non-significant positive interaction with *KAR3* (the budding yeast counterpart to kinesin-14/Klp2) hints at a potential functional connection, warranting further exploration.

Moreover, our research identified new interaction partners of *NRP1* within the cell polarity establishment module, indicating significant interactions with genes crucial for timing and reaction-diffusion pathways. These interactions may shed light on the mechanisms underlying polarity

recovery in $\Delta bem1$ mutants. We hypothesize that deleting *NRP1* could regulate critical cell size before bud formation, facilitating progression to subsequent cell cycle stages in $\Delta bem1$ mutants. We envision that the mechanisms by which the RNA binding domain of Nrp1 can influence those processes are nuclear transport or/and nuclear import and export. These are biological processes that were also found to be related to *NRP1* predicted positive genetic interactors.

Overall, our research showcases that sometimes suppressors can act in modules unrelated to the initial disrupted module, exhibiting the complex interconnection nature of the genetic interaction maps, and that conserved protein domains, like the RBD, could be accompanied by functional conservation across related species.

Compensatory gene disruptions revert the fitness effects of multiple mutations restoring wild type modularity

Mutations that suppress the effects of gene loss are frequently found in genes that are functionally related to the disrupted gene [28, 29]. However, the highly interconnected nature of genetic interaction maps implies that the consequences of compensatory mutations can have effects that go beyond the initially perturbed module. Understanding these mechanisms is crucial for unraveling the mysteries of evolution and adaptation.

Chapter 5 tackles this challenge by examining genome-wide gene interactions after the emergence of a suppressor mutation in an evolutionary trajectory. Specifically, we focus on the genetic adaptations of a sick yeast mutant, initially triggered by the loss of a nearly essential gene, *BEM1*, that subsequently progresses through an evolutionary trajectory marked by the sequential inactivation of additional genes, namely *BEM3* and *NRP1*. We particularly investigate the alterations following the first adaptive gene deletion, *BEM3*, as this stage represents the most significant phenotypic changes during the evolutionary trajectory, particularly regarding the restored ability to carry out successful polarization events.

Our genome-wide analysis of genetic interactions revealed, firstly, that genetic interactions are highly contingent on specific experimental conditions yet exhibit significant functional overlap across different techniques. We compared the existing genetic interactors predicted from Synthetic Genetic Arrays (SGA) of *BEM1* with our predictions using SATAY. Most SATAY-predicted genetic interactors were classified as neutral from SGA predictions and vice versa. This result is notable because it suggests the existence of variable resolution in each of these techniques, mainly due to two factors: i) the way that fitness is measured and ii) the experimental artifacts of each technique, such as the likely acquisition of secondary mutations in the case of SGA (due to the entire strain construction strategy [30]) that distort the true fitness effects of single and double gene deletions. Concerning fitness measurement, an important difference is the existence of a tangible reference in the case of SGA compared to SATAY, which relies on the assumption that most gene deletions are neutral to the population. This phenomenon has also been noted for the identification of essential genes for survival in budding yeast, which greatly differs across strain backgrounds and environments ([25, 31, 32]). However, none of these studies have shown how different strategies to identify essential genes (meiotic segregants, CRISPR-Cas9 and Transposon-based mutagenesis screens) differ between them.

Furthermore, our analysis shows that suppressor mutations predominantly induce sign epistasis changes, indicating significant genetic rewiring accompanying phenotypic changes. Based on this finding, we propose that while the suppressor mutation may be associated with the initial functional defect, the consequences of the suppressor mutation on the genetic interaction

network are notable and widespread throughout the entire network. In addition, sign epistasis in the context of evolutionary adaptation has been known to reduce the number of accessible mutational trajectories under strong selection [16, 18] which is an insight that implies that strong genetic constraint on the selective accessibility of trajectories to high fitness genotypes may exist and therefore evolution, in these cases, may be reproducible and even predictable [18]. This reasoning may theoretically explain why a loss-of-function mutation on the gene *NRP1* causes the population to evolve to a higher fitness state in at least half of the evolutionary lines. We confirm that compensatory evolution induces widespread rewiring and underscores the critical role of the genetic background in understanding the functional organization of the cell. This understanding is crucial for designing cancer therapies for tumor removal that target specific genes or functions to block tumor growth [33, 34].

Furthermore, our findings highlight that the effectiveness of loss-of-function mutations in *BEM3* to suppress the $\Delta bem1$ phenotype relies on genes unrelated to cell polarity, such as glycosylation, cell wall, vesicle traffic, and chromosome segregation. This suggests that while the suppressor mutation is associated with the same module as the initial functional defect, the epistatic constraints imposed on the cell indicate that further compensatory steps may not necessarily be linked to the impaired function. This observation may elucidate why the loss-of-function mutation in *NRP1*, a gene not assigned to the bud emergence module, represents a significant compensatory step in this evolutionary trajectory [4].

6

On the other hand, our functional enrichment analysis revealed distinct patterns in the genetic interactors of crippled versus recovered mutants, indicating a complex interplay of genetic interactions and functional modules in response to perturbations. For instance, our results suggest that in the genetic interaction network following the suppressor mutation, there is a clear distinction between functions whose alteration may have a positive or negative effect on fitness, in contrast to the genetic interaction network of the initial deleterious mutation. Based on this finding, we propose that predicting compensatory mutations for crippled mutants, resulting from the loss of a nearly essential gene, is more complex compared to already recovered mutants. Using the metaphor of the fitness landscape (the relationship between genotypic space and fitness), predicting the next step is comparatively "easier" when the fitness state is near a peak. In such scenarios, there are fewer alternatives towards reaching a peak, making the trajectory more predictable. However, when the fitness state is much further from a peak, numerous alternative paths towards optimization can emerge, complicating the prediction of the next step. Our SATAY predictions of compensatory mutations are more accurate for the case of the deleterious mutation. However, factors like strain background, media, and statistical significance can affect the interpretation of this result.

Overall, we propose that the genetic rewiring observed in sick yeast mutants undergoing compensatory evolution leads to a major reversion of the fitness effects of mutations, accompanied by a transition towards a more modular scenario. In this state, it is theoretically "clearer" which functions are beneficial or detrimental to the fitness of the population. Some functions would have a distinct and predictable impact on the cell's fitness, resembling behavior in normal physiological conditions. However, the extent to which compensatory mutations lead to similar effects remains an intriguing question for future research.

Bibliography

1. Doebeli, M. & Ispolatov, I. Chaos and unpredictability in evolution. *Evolution* **68**, 1365–1373 (2014).
2. Mas, A., Lagadeuc, Y. & Vandenkoornhuise, P. Reflections on the predictability of evolution: toward a conceptual framework. *Isience* **23** (2020).
3. Wortel, M. T. *et al.* Towards evolutionary predictions: Current promises and challenges. *Evolutionary applications* **16**, 3–21 (2023).
4. Laan, L., Koschwanetz, J. H. & Murray, A. W. Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *Elife* **4**, e09638 (2015).
5. Lind, P. A., Libby, E., Herzog, J. & Rainey, P. B. Predicting mutational routes to new adaptive phenotypes. *Elife* **8**, e38822 (2019).
6. Szendro, I. G., Schenk, M. F., Franke, J., Krug, J. & Visser, J. A. G. M. d. Quantitative analyses of empirical fitness landscapes. en. **2013**. Publisher: IOP Publishing, P01005. ISSN: 1742-5468. <https://doi.org/10.1088/1742-5468/2013/01/p01005> (2021) (Jan. 2013).
7. De Visser, J. A. G., Elena, S. F., Fragata, I. & Matuszewski, S. The utility of fitness landscapes and big data for predicting evolution. *Heredity* **121**, 401–405 (2018).
8. Johnson, M. S., Reddy, G. & Desai, M. M. Epistasis and evolution: recent advances and an outlook for prediction. *BMC biology* **21**, 120 (2023).
9. Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. *Nature* **490**, 535–538 (2012).
10. Zhou, J. *et al.* Higher-order epistasis and phenotypic prediction. *Proceedings of the National Academy of Sciences* **119**, e2204233119 (2022).
11. Papp, B., Pál, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194–197 (2003).
12. Conant, G. C. & Wolfe, K. H. Turning a hobby into a job: how duplicated genes find new functions. *Nature Reviews Genetics* **9**, 938–950 (2008).
13. Diepeveen, E. T., de la Cruz, L. I. & Laan, L. Evolutionary dynamics in the fungal polarization network, a mechanistic perspective. *Biophysical Reviews* **9**, 375–387 (2017).
14. Lamas, I., Merlini, L., Vještica, A., Vincenzetti, V. & Martin, S. G. Optogenetics reveals Cdc42 local activation by scaffold-mediated positive feedback and Ras GTPase. *PLoS biology* **18**, e3000600 (2020).
15. Martin, S. G. & Arkowitz, R. A. Cell polarization in budding and fission yeasts. *FEMS microbiology reviews* **38**, 228–253 (2014).
16. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).

17. Gorter, F. A., Aarts, M. G., Zwaan, B. J. & de Visser, J. A. G. Local fitness landscapes predict yeast evolutionary dynamics in directionally changing environments. *Genetics* **208**, 307–322 (2018).
18. De Visser, J. A. G. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**, 480–490 (2014).
19. Butland, G. *et al.* eSGA: E. coli synthetic genetic array analysis. *Nature methods* **5**, 789–795 (2008).
20. Kogenaru, M., de Vos, M. G. & Tans, S. J. Revealing evolutionary pathways by fitness landscape reconstruction. *Critical reviews in biochemistry and molecular biology* **44**, 169–174 (2009).
21. Usaj, M. *et al.* TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. *G3 Genes | Genomes | Genetics* **7**, 1539–1549. ISSN: 2160-1836. <https://doi.org/10.1534/g3.117.040220> (2022) (May 2017).
22. Tong, A. H. Y. *et al.* Global mapping of the yeast genetic interaction network. *science* **303**, 808–813 (2004).
23. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *science* **285**, 901–906 (1999).
24. Bosch-Guiteras, N. & van Leeuwen, J. Exploring conditional gene essentiality through systems genetics approaches in yeast. *Current Opinion in Genetics & Development* **76**, 101963 (2022).
25. Van Leeuwen, J. *et al.* Systematic analysis of bypass suppression of essential genes. *Molecular systems biology* **16**, e9828 (2020).
26. Baryshnikova, A. *et al.* Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature methods* **7**, 1017–1024 (2010).
27. Kingma, E., Diepeveen, E. T., de la Cruz, L. I. & Laan, L. Pleiotropy drives evolutionary repair of the responsiveness of polarized cell growth to environmental cues. *Frontiers in Microbiology* **14** (2023).
28. Helsen, J. *et al.* Gene loss predictably drives evolutionary adaptation. *Molecular Biology and Evolution* **37**, 2989–3002 (2020).
29. Szamecz, B. *et al.* The genomic landscape of compensatory evolution. *PLoS biology* **12**, e1001935 (2014).
30. Tong, A. H. Y. & Boone, C. Synthetic genetic array analysis in *Saccharomyces cerevisiae*. *eng. Methods in Molecular Biology (Clifton, N.J.)* **313**, 171–192. ISSN: 1064-3745 (2006).
31. Bajić, D., Moreno-Fenoll, C. & Poyatos, J. F. Rewiring of genetic networks in response to modification of genetic background. *Genome Biology and Evolution* **6**, 3267–3280 (2014).
32. Kinsler, G., Geiler-Samerotte, K. & Petrov, D. A. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *Elife* **9**, e61271 (2020).
33. Roth, J. A. & Cristiano, R. J. Gene therapy for cancer: what have we done and where are we going? *Journal of the National Cancer Institute* **89**, 21–39 (1997).
34. Cross, D. & Burmester, J. K. Gene therapy for cancer treatment: past, present and future. *Clinical medicine & research* **4**, 218–227 (2006).

Acknowledgements

I would like to start my acknowledgments by thanking life for all the moments and lessons that, over time, have helped guide my path toward this PhD, which I have enjoyed immensely—and also, why not admit it, suffered through. I think it would be unfair to only mention the incredible people I've met and who have contributed to me during these years of my doctorate, without also mentioning those people, moments, and circumstances that made it possible for me to begin my exploration in the biological sciences, starting from Physics.

I'd like to share what, for me, was the seed of my curiosity for biology, even before I started studying Physics at university. Life circumstances meant that my brother had a motorcycle accident while on his way home from work in Cuba. As a result, my brother had to undergo brain surgery to save his life. Thanks to everyone who intervened, my brother is healthy and safe, with no lasting consequences from the accident. However, this particular situation made me dive deeper into brain anatomy and physiology because I wanted to understand all the medical jargon surrounding this topic. Well, long story short, I fell in love with biology and medicine, and the more I read, the more I liked it, so I made it my life's mission to study or apply my studies to these subjects. A few years later, it was time for me to start university, and I studied Physics because I felt an equal attraction to it, but I still had the "biology and medicine bug" inside me. My mother is also a physicist, and I always remember her passion for science when explaining anything to me, regardless of my age.

As part of my studies in Cuba, I learned about the Neuroscience Center of Cuba, where I did my graduation thesis. I want to thank my supervisor, Nelson Trujillo Barreto, for teaching me and educating me on the scientific method and how to work interdisciplinarity.

This center also brought along something else—it's where I met the love of my life, the person who stands by my side now: Jose Carlos Urrea Llanusa. Together with him, I have learned what it means to fall, get back up, and keep going. He is my other half; we've gone through so much, so intensely, that I don't think I'd know how to live without him. I love you so much, and I tell you this often—I love you and will always love you.

Meeting him also gave me the chance to come to Delft, and thanks to luck and other mysteries, I studied my Master's in Applied Physics with an excellence scholarship awarded by Delft University. During my Master's, by the chance of fate, I had the opportunity to meet Liedewij Laan (thanks to Christophe Danelon, who put me in contact with her). And this is where this part of the story begins and where many of you may relate. Liedewij welcomed me on that first day with great enthusiasm and, I'd say, a bit of innocence. She explained the experimental system and the main research questions in the lab with such passion that I was hooked and full of ideas after that first conversation.

Liedewij, thank you so much for accepting me into your group and continuing to accept me through the different stages that followed. Thank you for teaching me how to be a scientist. Thank you for your critiques, questions, consistency, and patience, and for letting me explore areas where you might not have felt so confident but trusted me to venture into them. Maybe without knowing it, Liedewij, you helped Jose and me when we needed it most to keep living in

the Netherlands. So thank you so much for that, too. Liedewij, thank you also for giving me the opportunity to do my Master's thesis under the supervision of Werner Daalman, who taught me rigor, precision, and how to design experiments, who showed me how to run the first controls in the lab, and who taught me that good software can indeed be written in Matlab, with amazing GUIs. Werner Daalman, I admire you so much—thank you for teaching me so much and being such a wonderful person.

During my first years in the Master's, I met one of my best friends, Fayeze Shamsi. Fayeze, thank you so much for teaching me that a true friend can also be someone from the other side of the world. You taught me values like loyalty and that sometimes you need to open your mind and let go of prejudices to put yourself in the other person's shoes. Fayeze, you are such a special person, and I care for you deeply. Another person I met early on in the lab who has meant a lot to me is Enzo Kingma. Enzo, I've had such a great time with you! Thanks for being my fun partner at the borrels, dominos, table football, thanks for teaming up with me, under the cryptic name ".cu". You were my peer in the lab's Caribbean spirit, even if you didn't lift a finger when it came to dancing, but that's okay—I forgive you for all the fun we had. During the PhD, you made me grow as a scientist, thanks for pushing me out of my comfort zone and always being there to question the unquestionable, you taught me a lot. And I'm also so happy that you're my friend because that's how I see you. You have a friend in me forever.

6 Other special people who have marked my PhD and my life are Eveline, Shazia, Gregory, Wessel, Floor, Felix, Thomas, Ilse, Miranda, Marit, Els, Ingmar. Shazia, thank you for being such a great support, alongside Fayeze, when I became a mom. Eveline, thank you for planting the seed for SATAY development in our lab; I admire you for having the courage to move on with your life outside academia after such a long journey. Gregory, thank you for teaching me how to lead a collaborative software project, from which I learned much of what I know today about open software development and which marked my new journey toward training researchers with tools and skills that foster open science practices. Wessel, Floor, Ingmar, and Thomas, thank you for contributing valuable knowledge to the SATAY project. Many things in this thesis are thanks to your work on this project; thank you!

Also, many thanks to Sophie and Christine for bringing that German vibe to the lab, where order and rigor are very important :) Special mention to Sophie for dedicating so much time to the organization of the lab!

Of course, I can't forget to mention Ramon, who is our lab technician and also the name of one of my son's dragons :) Because who wouldn't want to be called Ramon, hahaha. Thank you so much, Ramon, for being so kind and such a good person. Thanks for helping so many people and surprising me with almost 50 balloons at my desk on my birthday—I will never forget it. Without you, our lab wouldn't be as cool. Keep being that special person that everyone needs. I wish you the best in your new role at BN.

I also want to thank Esengul for her good vibes and for helping so many students in the lab and department. Thanks also to Caspar for the great and interesting conversations, to Valentina for sharing topics with me that aren't often talked about in academia but take courage to face, and to Marieke and Nynke, thank you so much for keeping the lab filled with great vibes and ideas. Also, thank you, Frank, for your special charisma and teaching us how to play Golfito well!

And of course, there are many people outside the lab who have meant a lot to me during my years at BN, starting chronologically: Jose Ocampo (Pepe), Filip, Eve, Diego, Mathia Arens, Sadaf, Jeremy, Sacha, Fridtjof, Erwin, Ana, Daniel, Tanja, Nils, Catia, Aleksandre.

Pepe!! My buddy from the Master's and dear friend!! What a joy it was to meet you on the first day of the master's at the barbecue. I could speak Spanish with someone in the master's. We shared so many beautiful and memorable moments, and you were the door to meet all the Mexicans we know, and I always feel very proud to be "the friend of Pepe". You're special and humble, and I admire you a lot. I hope I can visit you in Chiapas one day and attend one of your classes!!!

Filip! How could I not mention you? You were the first Dutchman we tamed as a Latino, hahahaha! So many stories with you! I'll never forget how you invited us to Leeuwarden with a few others, and ten more were invited by the Mexicans, hahaha, and you didn't say no—on the contrary, you opened your house without hesitation. Special mention to your parents and sister Laura for welcoming us and being so kind to everyone. You are very special, and I consider you a dear friend.

Eve, thank you so much for your pleasant conversations in the hallway and being one of the first people with whom I found a friend in BN.

Diego, thank you so much for being so kind to me! You helped me a lot with the flow cytometer (FACS), and I'm very grateful because you always gave me your time and smiled, no matter how busy or stressed you might have been! I really appreciate it!

Mathia, thank you for fighting for life until the very end and showing us that life should be lived intensely until the last minute. You are a shining star for everyone. Rest in peace.

Sadaf and Catia have been our adoptees, hahaha. Sadaf and Catia, my two dear friends, thank you for giving me your friendship and for having so much fun together. Plus, you are both amazing at table football, and everyone wanted to be part of your team, hahaha. I love you both so much!

I would also like to mention two people from BN who helped me a lot during my PhD with using and troubleshooting the microscope: Sacha and Jeremy. Thank you so much. Sacha, thank you also for the great conversations and connection with Cuba. I remember your phrase: "No pasara!" I hope we can see each other again and that things are going well in your new jobs.

Thanks also to Fridtjof and Erwin for starting a collaboration that became one of the most extended projects of my PhD, but from which I learned a lot. I learned that making mistakes and getting things wrong is part of an exploration and that sometimes, big ideas require significant sacrifices. I also learned that in collaborations, patience is a great weapon that makes you more resilient. Thank you for, perhaps without intending to, teaching me so much.

To Ana and Daniel, for being the ones I could chat with in my Cuban slang in the hallway. Even though you're not Cuban, we could connect 100%. Thank you for letting me release my energy, speaking as I do at home, and being such inspiring colleagues. I also appreciate your dedication and perseverance in continuing in the scientific field. I love you both and wish you all the best in your future careers.

I'm also grateful to Tanja and Nils for being so kind and thoughtful to me and my family. You always remembered me at every party you organized and asked me how I was doing regarding motherhood. I wish you all the best on your return to Germany, and I hope we can also meet again! Also, thanks to Aleksandre for the nice and funny conversations in the hallway; I admire your easy-going spirit and hard work. I wish you the best in your flourishing career.

I cannot forget to mention Joyce, our beloved Joyce. We all know her, and she is one of the most important people at BN. She ensures that we work in a clean and beautiful environment every day. I consider her part of my family here in the Netherlands. She's incredible, with a huge

heart and an angelic presence.

To my lifelong friends, my motherhood companions, Jenny and Febe, who always give me the strength to keep going and who are an example of courage and love to me.

To my dear friend Heather, whom I met during this PhD journey and who is here to stay, thank you for your beautiful friendship.

To Casandra and David, thank you for always being there when we needed you, and for your constant support.

Thanks also to Jan van der Heul, a colleague and friend at my current job at 4TU.ResearchData. He is a truly special person who will immortalize my defense day through his photography. Thank you for this special gift!

Thanks also to the members of this thesis committee who kindly dedicated their time to read my thesis and provide valuable feedback. I genuinely thank you.

Lastly, I want to say a few words to my dear and special family. Mami, I love you so, so much. Thank you for teaching me to fly and to fight for what I want with perseverance and courage. You have been an incredibly brave woman, and I owe everything I know to you. Papa, a big kiss. You are another star that protects me. Thank you for being so good and smiling at life, even when it didn't smile back at you. You taught me resilience and how to keep moving forward, even when facing seemingly unbreakable walls. To my brother, thank you for always being there!!! My bro, my "torbo", my big brother! You are very special to me, and I love you so much. My grandma Chicha is another of my stars, perhaps the biggest one—you must be here too! For being such a wonderful person, so full of magic and sensitivity. You inspired me to become a mother of more than one! I loved seeing you teach and be so beloved by so many!

To the family that love brought to me, Pedri, Susy, and Haydecita, I love you all so much! Thank you for taking such good care of us and teaching me many lessons of love and sacrifice.

Finally, to my children, Pablo and Mina, thank you for teaching me what it truly means to live and showing me the purest form of love and the most terrifying fears. You are my life and my greatest joy.

List of Publications

Peer-reviewed publications

4. Enzo Kingma, F Dolsma, **Leila Iñigo De La Cruz** and Liedewij Laan *Saturated Transposon Analysis in Yeast as a One-step Method to Quantify the Fitness Effects of Gene Disruptions on a Genome-Wide Scale*, bioRxiv,2023.(Accepted in Plos One in October,2024)
3. Fridtjof Brauns, **Leila Iñigo de la Cruz**, Werner K.-G. Daalman, Ilse de Bruin, Jacob Halatek, Liedewij Laan and Erwin Frey, *Redundancy and the role of protein copy numbers in the cell polarization machinery of budding yeast*, Nature Communications,14,6504 (2023).
2. Enzo Kingma, Eveline T. Diepeveen, **Leila Iñigo de la Cruz** and Liedewij Laan, *Pleiotropy drives evolutionary repair of the responsiveness of polarized cell growth in Saccharomyces cerevisiae to environmental cues.*, Frontiers Microbiology, Sec. Microbial Physiology and Metabolism, Volume 14,2023.
1. Eveline T. Diepeveen, **Leila Iñigo de la Cruz**, and Liedewij Laan, *Evolutionary dynamics in the fungal polarization network, a mechanistic perspective.*,Biophysical reviews 9,375–387 (2017)

Open access research software

1. **Leila Iñigo de la Cruz**, van Beek, G., and Kok, M., *Transposonmapper (Version 1.1.5) [Computer software]*, Zenodo,2022.

Curriculum Vitæ

Leila Margarita IÑIGO DE LA CRUZ

30-07-1990 Born in Havana, Cuba.

Education

2008-2013 Summa Cum Laude Diploma in Physics Engineering
Facultad de Fisica, Universidad de la Habana

Thesis: Bayesian model averaging to tackle the inverse problem of the Electroencephalogram

Promotor: Prof.dr. Nelson Trujillo Barreto

2015-2017 MSc Applied Physics, Applied Science Faculty
Delft University of Technology

Thesis: The role of Cla4 in the establishment of cell polarity in budding yeast

Promotor: dr. Ir L. Laan

2018-2023 PhD. in Biophysics
Delft University of Technology
Faculty of Applied Sciences, Department of Bionanoscience

Thesis: Unlocking the hidden dance of cellular resilience: Exploring the evolutionary adaptation of the cell polarity machinery in *S.cerevisiae*

Promotor: dr.Ir. L. Laan

Promotor: Prof.dr. M. Dogterom

Awards

2015 Holder of a DRI Full Excellence Scholarship to study the MSc Applied Physics at TU Delft.

2021 Convergence Health and Technology Open Research Award from Convergence Alliance, Rotterdam R.I.O.T. Science Club, Open Science Community Rotterdam, and Open Science Community Delft