Exploiting Genetic Interactions in Metabolic Engineering

Cachera, Paul Pierre-Yves Jean

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

Exploiting Genetic Interactions in Metabolic Engineering

PhD Thesis

Paul Cachera

Supervisor: Michael Krogh Jensen, Co-supervisor: Uffe Hasbro Mortensen

The Novo Nordisk Center for Biosustainability
Technical University of Denmark
September 2023

Preface

This thesis serves as partial fulfillment of the requirements of obtaining a PhD degree at the Technical University of Denmark, PhD school of Novo Nordisk Center for Biosustainability. The work presented was carried out in the time period between September 2020 and September 2023 at the Novo Nordisk Foundation Center for Biosustainability and at the Department of Bioengineering and Biomedicine at the Technical University of Denmark, Kongens Lyngby.

The work was supervised by Senior Researcher Michael Krogh Jensen as a main supervisor and Professor Uffe Hasbro Mortensen as a co-supervisor. Senior Researcher Nikolaus Sonnenschein assumed the main supervision of this project from September 2020 to July 2022 before resigning from his researcher position at DTU Bioengineering. Funding was provided by the Novo Nordisk Foundation through the Copenhagen Bioscience PhD Programme (grant number NNF19SA0035438).

Paul Cachera, September 2023.

Acknowledgments

This work and its author have benefited from the input and support of a great deal of generous, talented, smart and patient people who deserve a big round of applause.

The first round goes to my three supervisors, Michael K. Jensen, Uffe H. Mortensen and Nikolaus Sonneschein for their scientific guidance and help designing my project. I have greatly benefited from their complementary expertise and management styles. I also thank them for cheering me up when Science was unkind.

My office and bench mates and fellow SBTY group members also deserve a big thank you for maintaining a pleasant working atmosphere and complying with my rigorous zero-noise policy. I thank Tomas Strucko from the Mortensen Group for his experimental tips, dark humour and invincible optimism.

I thank Helén Olsson who trusted me in overtaking CRI-SPA, the method she had developped over her entire PhD, by sharing her data, protocols, strains and plasmids. This made the knowledge transfer for CRI-SPA flowless.

I thank Pr. David Luo at the Shenzhen Institute of Advanced Technology for the warm welcome I received from him and his lab during my external stay.

I thank the talented students Mads Liengaard, Andreas Thorning Røpke and Nikolaj Can Kurt who conducted their research projects under my supervision and who contributed to some of the work presented here. Reasoning with them and seeing them learn and gain autonomy was one of the most rewarding part of my PhD.

I thank the PhD School and in particular Rebeca Thostrup and Morten Nørholm for accompanying PhD students throughout their studies and being attentive to their well-being.

I thank the Novo Nordisk Foundation for allowing me to take part in the Copenhagen Bioscience Program and to fund the best possible research conditions at the Center for Biosustainability. Joining the program and moving to Copenhagen has undoubtedly been one of the most mentally stimulating and gratifying periods of my life. I also thanks Amelia Green and Moreno Papetti for their professionalism in coordinating the program.

Finally, I thank my parents and my family to whom I owe everything else. In particular, I thank my grandmother for her indefatigable curiosity concerning my research subject.

Contents

0.1 Popular science summary

By assembling parts with given properties, engineers are able to build tools and devices of increasing complexity. For example, a simple Ikea table is assembled from pieces of wood, screws and bolts with precisely designed dimensions, diameters and thread angles. The properties of these parts do not change upon assembly, allowing for a predictable outcome: a table.

This simple reductionist principle is also central to synthetic biology, the field interested in engineering living organisms. In synthetic biology, DNA sequences encoding biological parts found in the living world are copied, assembled and reintroduced in other organisms to engineer new biological functions. For example, the enzymes taking part into a metabolic pathway synthesising a plant pharmaceutical ingredient can be expressed in the yeast *Saccharomyces cerevisiae* allowing for its synthesis by cultivating the modified yeast in a bioreactor.

The applicability of the reductionist approach underlying traditional engineering disciplines is however limited in synthetic biology. Indeed, the properties of biological parts are often dependent on the context of the host they are expressed in. For example, all parts expressed in a living host are synthesised from a family of building blocks (e.g. amino acids, lipids, sugars) which availability is not infinite. Cross-talk between parts might arise simply by competing for the same pool of resources. These interactions limit the predictability of designs and, in practice, many iterations of trial-and-error are often needed to engineer a synthetic biological mechanism.

In experimental genetics, the unexpected outcome arising from the combination of two genetic modifications is known as a *genetic interaction*. Genetic interactions can be complex and often elude our understanding, even in well studied model organisms. Yet, in rare instances, they produce a surprisingly strong phenotype hinting that they may be exploited in an engineering context.

The motivation of this thesis is to turn interactions into a source of improvement for a desired biological function. Indeed, interactions between a metabolic pathway and a modification of a host gene may lead to higher product synthesis. In this thesis, I first attempt to predict interactions with machine learning. Failing to do so, I turn to a high throughput screening to identify positive interactions between a metabolic pathway and all non-essential genes in *S. cerevisiae*. More specifically, we use CRI-SPA, a new high throughput gene delivery method, to deliver a metabolic pathway in all the strains of either the Yeast Knock Out or the Yeast Over-expression Libraries. This systematically tests for the presence of an interaction between a gene (its absence or over-expression) and the metabolic pathway which might improve its yield. We show that this method can identify positive interactions in two case studies improving the synthesis of the plant pigment betaxanthin or that of the platform chemical *cis-cis*-Muconic acid.

Altogether, this work shows that genetic interactions can be used to improve a desired engineered trait in synthetic biology and advocate for a shift from its original reductionist dogma.

0.2 Dansk Resumé

Ved at samle komponenter med specifikke egenskaber, er ingeniører i stand til at konstruere værktøjer og apparater af stigende kompleksitet. For eksempel samles et simpelt Ikea-bord af træstykker, skruer og bolte med nøjagtigt designede dimensioner, diametre og gevindvinkler. Egenskaberne ved disse komponenter ændrer sig ikke ved konstruktionen, hvilket muliggør et forudsigeligt resultat: et bord.
Dette simple reduktionistiske princip er også centralt i syntetisk biologi, et felt der bestræber sig på at designe levende organismer. I syntetisk biologi kopieres DNA-sekvenser, der koder for biologiske komponenter, der findes i naturen, som samles og indsættes i andre organismer for at designe nye biologiske funktioner. For eksempel kan enzymer, der deltager i en metabolisk vej til syntese af et farmaceutisk stof fra planter, udtrykkes i gæren *Saccharomyces cerevisiae*, hvilket muliggør dens syntese ved dyrkning af den genmodificerede gær i en bioreaktor.

Anvendeligheden af den reduktionistiske tilgang er dog begrænset i syntetisk biologi. Egenskaberne ved biologiske komponenter afhænger ofte af sammenhængen i den værtsorganisme de udtrykkes i. For eksempel er alle komponenter, der udtrykkes i en levende vært, syntetiseret fra en familie af byggesten (f.eks. aminosyrer, lipider, sukkerstoffer), hvis tilgængelighed ikke er uendelig. Interferens mellem dele kan opståved konkurrence om den samme ressourcepulje. Disse interaktioner begrænser forudsigeligheden af design i syntetisk biologi. I praksis er der ofte et behov for mange iterationer af forsøg for at designe en funktionel syntetisk biologisk mekanisme.

I eksperimentel genetik kaldes det uventede resultat, der opstår som følge af kombinationen af to genetiske ændringer, for en genetisk interaktion. Genetiske interaktioner kan være komplekse og er ofte udenfor vores forståelse, selv i velundersøgte modelorganismer. Dog kan de i sjældne tilfælde producere en overraskende stærk fænotype, hvilket kan antyde, at de har potentiale for at blive udnyttet til et specifikt formål.

Motivationen for denne afhandling er at omdanne interaktioner til forbedring af en ønsket biologisk funktion. For eksempel kan interaktioner mellem en metabolisk vej og en ændring af et værtsgen føre til øget produktionskapacitet. I denne afhandling forsøger jeg først at forudsige interaktioner ved hjælp af machine learning. Når dette ikke lykkes, anvender jeg en high-throughput screeningsmetode til at identificere positive interaktioner mellem en metabolisk vej og alle ikke-væsentlige gener i *S. cerevisiae*. Mere specifikt bruger vi CRI-SPA, en ny high-throughput metode til levering af generfor en metabolisk vej til alle stammer i enten yeast knock out eller yeast over-expression biblioteker. Dette tester systematisk tilstedeværelsen af en interaktion mellem et gen (dets fravær eller overudtryk) og den metaboliske vej, der forbedrer dets udbytte. Vi viser, at denne metode kan identificere positive interaktioner i to case-studier, der forbedrer syntesen af plantepigmentet betaxanthin eller kemikaliet cis-cis-Muconic acid.

Alt i alt viser dette arbejde, at genetiske interaktioner kan bruges til at forbedre en ønsket egenskab inden for syntetisk biologi og argumenterer for en ændring i dets oprindelige reduktionistiske dogma.

0.3 Abstract

Life has evolved a marvelous diversity of mechanisms to conserve energy, move, receive, process and store information, cope with stress, etc. The engineering and repurposing of these functions would constitute a steady source of formidable technological disruption helping humanity overcome the critical climate, health and agricultural challenges it is facing.

With the rise of DNA editing, writing and sequencing, biologists have become able to manipulate DNA sequences and insert them in living organisms to implement new biological mechanisms. In the early 2000's, synthetic biology emerged as the scientific field rationalising this endeavour. Taking inspiration from the successes of other engineering disciplines, synthetic biology championed the use of a reductionist approach according to which natural biological systems could be deconstructed into modular parts. These parts, characterised by input-outputs properties, could, in principle, be predictably reassembled to create new synthetic functions of arbitrary complexity. Yet, close to a quarter of century later, the rational design of biological functions has not materialised and engineering efforts still heavily rely on the trial-and-error process of the Design-Built-Test-Learn cycle. So why do designs so often fail to produce the expected behaviours?

"The action of natural selection has often been compared to that of an engineer. This, however, does not seem to be a suitable comparison. [...] One would have to say that natural selection does not work as an engineer works. It works like a tinkerer" - *Francois Jacob, 1977*

The predictability of rationally assembled parts is plagued by unexpected **genetic interactions** taking place between parts and with the genetic context of the host they are introduced in. The effects of genetic interactions on genotype-phenotype relationships are well documented but the mechanisms behind them are rarely understood. In the case of synthetic biology, the ubiquity of genetic interactions advocates for a shift in its reductionist paradigm. The field needs to rethink its approach and possibly take inspiration from evolution, "the master tinkerer" (Jacob, 1977).

As we will see in this introduction, genetic interactions can produce a surprisingly beneficial effect on a phenotype. In this thesis, I attempt to develop a rational framework to harness these beneficial interactions in an engineering context. More specifically, the aim is to exploit interactions improving the synthesis ability of a microbial cell factory. Since interactions are rare and spread in a large combinatorial space **(Box 1, section 1.4)**, their identification requires the mean to either predict them with a computational tool or to experimentally search for them with a high throughput screen.

The experimental work presented in the thesis is split in three chapters. In Chapter 2, I attempt to employ machine learning to predict genetic interactions. Given the poor predictions obtained in this effort, I turned to an experimental screen quantifying the effects of individual genetic perturbations on the synthesis of a product. More specifically, Chapter 3 describes the development of a new screening method, CRI-SPA, enabling the systematic genome editing of yeast libraries. CRI-SPA is employed to deliver a metabolic pathway in the Yeast Knock-Out library to test for interactions between each gene and the pathway. In its first implementation, CRI-SPA uses the biosynthesis of betaxanthins, a family of colorful plant pigments, as a case study. The reliance on betaxanthins' color to readout a productivity metric limits its application to a handful of colourful or fluorescent compounds. In Chapter 4, CRI-SPA is adapted to work with a biosensor linking the synthesis of the inconspicuous chemical *cis-cis*-muconic acid to a fluorescent read-out. Finally, the possibility to continue the search for higher order interactions arising when several modified gene loci are combined is discussed in Chapter 5.

Chapter 1

The potential of genetic interactions in metabolic engineering

1.1 Defining genetic interactions

The occurrence of interactions between genetic elements was discovered by early geneticists and their relevance in shaping the genotype-to-phenotype relationship is still a subject of intense study in numerous branches of modern genetics (Bank, 2022; Domingo et al., 2019). The concept is sometimes referred to as *epistasis* and has adopted different connotations in population genetics, evolutionary biology and molecular and systems biology (Phillips, 2008; Bank, 2022). The term epistasis was initially coined by William Bateson to describe the masking effect of a locus on another, preventing the appearance of an expected phenotype in an organism (Bateson, 1909). Confusingly, Fisher used a similar term, *epistacy*, to describe a deviation from the expected additive effects of two genetic changes on the phenotype of a population. In evolutionary genetics, epistasis describes the non-linear effects on fitness of combinations of genetic edits explaining the presence local maxima in the fitness landscape (Wright et al., 1932; Kauffman and Levin, 1987; Bank et al., 2016; Bakerlee et al., 2022).

In modern molecular biology, the term *genetic interaction* describes the occurrence of an unexpected phenotype arising from the combination of two or more genetic modifications which do not produce such an effect on their own. In this context, the individual modifications and their combinations are experimentally introduced in an identical genetic background and controlled laboratory conditions allowing the quantification of the difference between the expected and the observed phenotypes (Costanzo et al., 2019; Domingo et al., 2019; Mani et al., 2008). Thanks to this framework, interactions between sequence edits have been mapped at the base-pair (Puchta et al., 2016), amino acid (Pokusaeva et al., 2019), and gene level (Costanzo et al., 2016; Kuzmin et al., 2018).

This thesis is interested in this last definition, where genetic interactions are observed as a result of genetic perturbations introduced at the gene level (e.g. knock-out, over-expression) in a microbial cell. This notion of interaction is mechanistically vague as it must depict non-linearities arising at any biological layer linking a genotype to its phenotype (Domingo et al., 2019). As I review next, genetic interactions range from simple functional overlap between two gene orthologs, to cryptic interdependence in gene functions preventing the identification of a truly minimal genome.

1.2 Building the case for genetic interactions

The presence of interactions between genetic sequences is not surprising given their evolutionary origin. Genes are not engineered in isolation, rather they arise from pre-existing sequences in a genomic

ecosystem. Hence, gene functions might co-evolve, overlap, complement one-another and form networks which might be overseen by a reductionist depiction (Bergelson et al., 2021). For example, duplication and divergence is a classical model explaining how new genetic functions arise (Taylor and Raes, 2004). This process allows genes to specialise and yet maintain a degree of overlap. Importantly, *S. cerevisiae* underwent a whole genome duplication which, after a number of rearrangements and deletions, left 500 duplicated gene pairs (Wolfe, 2015). Interaction between a pair might arise because one duplicate is often sufficient to maintain cellular viability when the other duplicate is mutated. However, deletion of both duplicate leads to a more severe fitness defect, i.e. a negative interaction, and in the most extreme case cell death, referred to as synthetic lethality. This pattern is commonly observed for metabolic genes where an enzyme maintains enough side activity to catalyze the reaction of a mutated paralog, contributing to metabolic networks' robustness (Harrison et al., 2007). Interactions due to function overlap is not limited to metabolic networks. In another classical example, *TUB1* and *TUB3*, two alpha-tubulin isotypes are known for their synthetic lethality. The loss of *TUB3* only causes a modest fitness defect while deletion of both *TUB1* and *TUB3* is lethal. *TUB1* deletion is also lethal but can be rescued by *TUB3* overexpression (Schatz et al., 1986). Although the mutual rescue of the isotypes suggests a strong overlapping role, it was recently shown that they have evolved specific binding properties with other cytoskeleton proteins (Nsamba et al., 2021; Moore and Wethekam, 2021).

The existence of interactions between gene pairs has been most comprehensively catalogued by the Synthetic Gene Array method (see description of SGA in section 1.5.3 below). The SGA platform has enabled the combination of close to all possible pairs (23 million) of gene knock-outs (KO) in yeast and the quantification of their effects on fitness. This systematic analysis has revealed close to a million interactions, defined as a more extreme fitness than expected based on the fitness of the single mutants. The biological mechanisms behind the interactions are, in the majority of cases, not elucidated. However, a number of systems-level learnings can be extracted from this double mutant dataset. Importantly, interactions are not randomly spread within the set of double mutants. For example, some genes form interaction "hubs" and genes with a similar function tend to have correlated interaction profiles **(fig. 1.1b)**. Similarly, genes taking part in the same cellular function (i.e. protein complexes, metabolic or signalling pathways) tend to show more positive interactions ('within pathway model', Costanzo et al. 2016, **fig. 1.1a)** while negative interactions are particularly frequent between genes taking part in two redundant functions ('between pathway model', Fang et al. 2019, **fig. 1.1a)**. While the double mutant dataset generated by SGA maps genome-level intergenic interactions, it is important to note the occurrence of intragenic interactions taking place between single base-pair edits is also extremely well documented (Domingo et al., 2019).

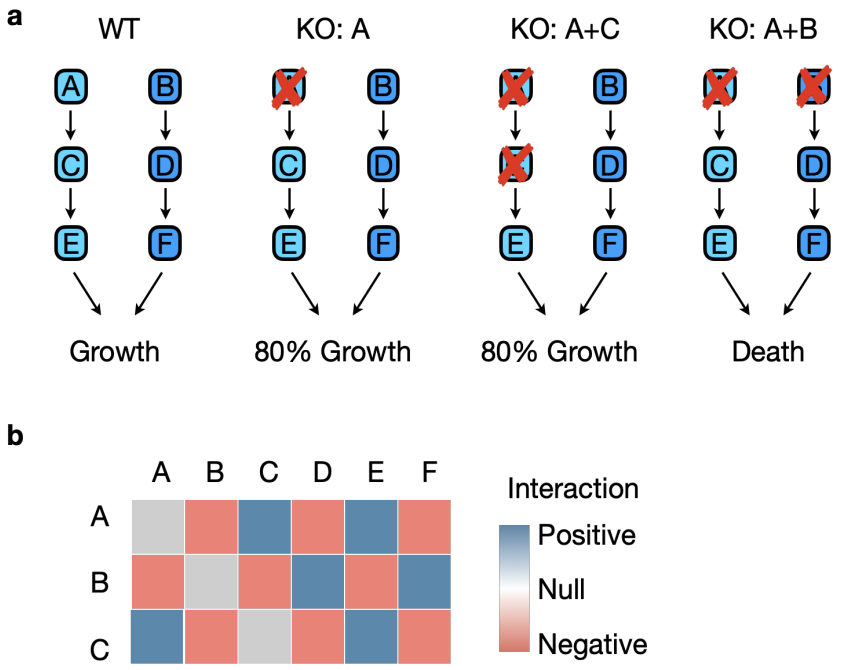


Figure 1.1: **Between/Within Pathway Interactions**
Toy example of 'Within' and 'Between Pathways' interactions. **a)** Left: Two alternative pathways contribute in the making of an essential metabolite. Center left: Deletion of A shuts down one of the pathways, leading to a 20% fitness defect. Center right: Additional deletion of C *within* the same pathway does not result in additional growth defect. A and C show a positive interaction. Right: Additional deletion of B which shuts down the alternative pathway, results in a negative interaction (here a synthetic lethal). **b)** Schematics of the resulting interaction profiles created by the toy example. The profile of A and C, which take part in the same pathway are highly correlated. (Panel a adapted from Fang et al. 2019)

The existence of conditionally essential genes is another compelling evidence that the function of certain genes is influenced by a particular genetic context. Conditional essentiality occurs when a gene is essential only in a given genetic background. For example, although S288c and Σ1278b are two closely related yeast strains, 44 genes are only essential in Σ1278b and 13 genes are only essential in S288c (Dowell et al., 2010). Here, gene functions depend on other genomic modifiers and the complexity of these interactions is only starting to be appreciated (Hou et al., 2019).

The fact that the function of some genes depends on their genetic context is also apparent in genome reduction efforts. Top-down approaches progressively delete genomic regions reducing the genome to a minimal set of genes which still supports viability. Intriguingly, the essentiality of a remaining gene depends on the set of genes previously deleted. In *E. coli*, this idiosyncrasy results in an only partial overlap in the sets of genes isolated by different reduction studies (Mizoguchi et al., 2008; Xu et al., 2023). This illustrates that gene functions are intimately linked to a given background complicating the isolation of a truly minimal genome.

The interactions of biological functions and their dependence on a given context has proven to be a major obstacle to their engineering. Next, we describe how genetic interactions challenge the field of synthetic biology and advocate for a different approach which would use them to the advantage of an engineering task.

1.3 Genetic interactions in synthetic biology

The field of synthetic biology seeks to make the engineering of biology simpler and more predictable (Kitney and Freemont, 2012). Chiefly, synthetic biology deconstructs biological systems into smaller parts which can be characterised, optimized and reassembled. Akin electronic components, these parts, (e.g. promoters, terminators, RBSs) can be stored in registries (Canton et al., 2008) and assembled with the help of softwares to create biological mechanisms of great complexity (Jones et al., 2022). The analogy to electronics has been extended to the building biological functions mimicking their in silico counterparts. For example, genetic circuits could be wired to perform simple computations (Moon et al., 2012; Nielsen et al., 2016; Tamsir et al., 2011), inducible invertases could write digital-like data into DNA (Bonnet et al., 2012; Friedland et al., 2009) and light sensitive biosensors enabled transcriptional control of prokaryotic genes with light, enabling bacterial 'RGB vision' and 'edge detection' (Levskaya et al., 2009; Tabor et al., 2009; Fernandez-Rodriguez et al., 2017). The promise to make biology engineerable like electronics infused great enthusiasm into the field as a new source of disruptive technologies.

Ten years on, synthetic biology has matured beyond its original electronic metaphor and enabled undeniable academic successes (Meng and Ellis, 2020). However, in light of the original enthusiasm, its ability to create disruptive biotechnologies has arguably not lived up to expectations. Tellingly, only a handful of synthetic biology-enabled technologies have been marketed (Voigt, 2020) with some of them being the results of research long predating synthetic biology (Gross et al., 1989). Further, the publication of "complex" biological systems in high impact journals supports ironically that their assembly is still a technical feat and not a routine engineering practice. Evidently, cells remain difficult to engineer, more so than synthetic biology as a nascent field once suggested.

The idea that biological systems can be dismantled into orthogonal parts might hold true for the assembly of a limited number of very well characterised components (Pbad, Ptet, Plac, T7RNAP, GFP, etc., Bergelson et al. 2021). But these represent a minority of functional elements taking parts in the central dogma (transcription, translation) which has been thoroughly studied in model organisms. Yet, even in *S. cerevisiae*'s, a sixth of the gene pool remains uncharacterised (Pena-Castillo and Hughes, 2007; Xu et al., 2023) and most cellular mechanisms are still outside the reach of such rational design. Second, the rate of synthesis of energy and building blocks in a growing cell are limited and their allocation between energy generation and self-replication is balanced to achieve fastest growth (Scott and Hwa, 2023; Weiße et al., 2015). Heterologous parts introduced in the 'cellular economy' are likely to disrupt this balance. Because these pools are limited, interactions between heterologous parts (Gyorgy et al., 2015; Shopera et al., 2017; Qian et al., 2017), and with their host prevent the assembly of arbitrarily complex functions (Boo et al., 2019). In the worst case, heterologous parts create a metabolic burden and hence an evolutionary pressure on the host to eliminate them. Concerns about the limitations caused by the interactions between parts and their host were raised early on (de Lorenzo, 2011; Cardinale and Arkin, 2012) and have started to be addressed by a wave of "host aware" synthetic biology (Boo et al., 2019).

Broadly speaking, host-aware synthetic biology devises two major rational approaches to alleviate host:construct interactions (de Lorenzo, 2011; Cardinale and Arkin, 2012). The first strategy is to alleviate burden by controlling the expression of heterologous sequences with feedback regulation (Dahl et al., 2013; Siu et al., 2018; Segall-Shapiro et al., 2018; Ceroni et al., 2018; Barajas et al., 2022). For example, engineering a feedback responsive to a toxic intermediate can tune the expression of enzymes in heterologous pathways to reduce the accumulation of this intermediate (Dahl et al., 2013; Siu et al., 2018). In another example of elegant dynamic regulation, Ceroni et al. placed the expression of a burdening construct under the negative regulation of a 'burden responsive' promoter, adapting expression levels to the metabolic capacity of the host and achieving better yield (Ceroni et al., 2018).

The second strategy seeks to reduce host:construct interactions by isolating them from the host. For example, transcription and translation insulators have been successfully introduced in complex circuit designs to reduce crosstalk between parts and with the host. At the transcription level, bacterial sigma factors recognising different promoters can drive activation of several genes with little cross talk (Segall-Shapiro et al., 2014; Fernandez-Rodriguez et al., 2017). T7 phage polymerases driving transcription of their cognate promoters are also very popular to drive orthogonal gene expression in synthetic biology (Meyer et al., 2015; Costello and Badran, 2021). At the translational level, self-cleaving ribozymes can be added upstream of the RBS to cleave the 5' untranslated region and reduce its effects on translation initiation (Nielsen et al., 2016; Lou et al., 2012). This concept of orthogonality can be extended to translation with orthogonal antiShineDalgarno driven ribosomes (Rackham and Chin, 2005; Orelle et al., 2015; Darlington et al., 2018) and repurposed synonymous codons enabling the incorporation of noncanonical amino acids (Tang et al., 2022; Robertson et al., 2021). These advances have led prominent synthetic biologists to speculate that a full "orthogonal dogma" was on the horizon (Costello and Badran, 2021; Liu et al., 2018). Despite its exciting academic implications, an orthogonal dogma will likely be subject to the same metabolic constraints and therefore unlikely to abrogate genetic interactions.

1.4 The potential of genetic interactions in synthetic biology

Beyond questioning the pertinence of its 'orthogonal dogma', genetic interactions might also represent an untapped source of improvements for synthetic biology. The completeness of the SGA dataset gives us an idea of the properties of the interaction landscape (Costanzo et al., 2016). First, interactions are rare, representing only 4% of the 23 million pairs tested. Second, interactions can be both positive and negative, meaning that the fitness of the double mutant can be more extreme than expected based on the fitness of the single KO. Last and most important, the fittest gene pairs in the landscape are the results of interactions between genes which do not produce such an improvement on their own. This hints that interactions can be exploited within a strain engineering context to obtain synergistic effects on a desired phenotype.

Evidence supporting the importance of considering interactions in synthetic biology is accumulating. For example, the behaviour of genetic circuits have been reported to differ across different host strains and species (Chan et al., 2023; Khan et al., 2020; Tas et al., 2021). Strucko et al. observed a ten-fold improvement in vanillin-β-glucoside titers when expressing the same pathway in S288c compared to CEN.PK (Strucko et al., 2015). Further, the Zhao lab which pioneers the use of combinatorial methods for metabolic engineering, regularly reports synergistic combinations of edits which show little improvement when reintroduced on their own (Bao et al., 2018; Lian et al., 2017, 2019). For example, the concerted upregulation of *PDI1* (involved in disulfide-bond formation in the endoplasmic reticulum (ER) lumen), inhibition of *MNN9* (responsible for protein mannosylation in the ER) and KO of *PMR1* (involved in Ca^{2+} transport in the Golgi) was found to promote the expression and surface-display of a heterologous protein of interest. Nevertheless, none of these modifications increased the expression of this protein when tested on their own (Lian et al., 2017).

Since positive interactions are rare and spread in a large combinatorial space **(Box1)**, synthetic biologists need either computational tools to predict them or a high throughput experimental setup to screen the combinatorial space. The possibility of predicting genetic interactions with machine learning has been investigated in this thesis and is reported in Chapter 2. Since this predictive approach was unsuccessful, I next committed to an experimental approach to screen for interactions. This requires a methods to introduce genetic perturbations and screen their effects on a phenotype at high throughput. In the second part of this Introduction, I review some of the existing high throughput methods linking genotype to phenotype enabling the exploration of the genetic space.

Box1: Combinatorial explosion of the interaction space.

The number of possible combinations in the interaction space increases exponentially with the number of edits considered. This is known as combinatorial explosion. For example, if one wanted to change the state S of p positions drawn from a larger pool of G positions, the number of combinations N would be given by:

$$N = \frac{G!}{p!(G-p)!}S^p \tag{1.1}$$

To avoid the explosion of N, the size of G, p and S must be constrained. This trade-off is sometimes referred to as a balancing the 'breadth' (i.e. increasing G) and 'depth' (i.e. increasing p and/or S) of a screen (Pokusaeva et al., 2019; Bakerlee et al., 2022). As an illustration, SGA has covered most of all the possible double mutant combinations in yeast. This represents a space defined by p=2, $G \approx$ 6,500 (gene pool in *S. cerevisiae*), and S=1 (one change in state: KO) and which results in N $\approx$ (6500*6499/2)*1 $\approx$ 20 million combinations. While screening such a large number of strains represents an admirable technical prowess, the size of G constrains the increase of p (i.e. higher level combinations) and S (i.e. including more gene expression levels) which remain unexplored. More recently, Kuzmin et al. started venturing in the triple mutant space (i.e. increasing p to 3), generating 200,000 triple mutants (Kuzmin et al., 2018). Completing this triple mutant library would require the building of close to 46 billion strains! If the authors wanted to also test for gene overexpression (i.e. S=2), this number would be multiplied by eight.

1.5 Tools for exploring the genetic space

A wide variety of approaches leveraging new genetic tools, robotics, biosensing and advances in DNA synthesis and sequencing have been developed since the 2000s. Broadly speaking, genome-wide editing strategies can be classified based on the degree of control that they grant the user. At one end of the spectrum, strategies relying on spontaneous mutations such as adaptive laboratory evolution (ALE) are fully non targeted but are free of constraints on the number, nature and locations of genomic perturbations. CRISPR and recombineering strategies offer more control on the nature and locations of edits and are therefore referred to as semi-rational (Yilmaz et al., 2022). Finally, systematic genome-wide screens enabled by yeast libraries (Giaever et al., 2002; Yofe et al., 2016; Arita et al., 2021) and the SGA method offer the possibility to test all genetic modifications in the library separately. Instead of performing genome modifications on a strain harbouring a genetic feature of interest, the feature is combined at high throughput with features already present in a genetic library. Each modification in the library is separately tested providing the most comprehensive datasets (Costanzo et al., 2016; Kuzmin et al., 2018). However, the genomic perturbations (e.g. gene KO) and the genomic backgrounds in which the feature is tested are predefined by the choice of the library.

Here, we focus on ALE, genome-wide CRISPR editing and SGA in the context of genome space exploration. We compare their abilities to find strains with improved phenotypes and to generate informative data on the mechanistic causes of these improvements. These two abilities hit a tradeoff similar to the "exploration-exploitation" concept described in active learning. Active learning is the branch of machine learning interested in actively sampling the data space to train a model with the least amount of data. In active learning, the learning space is sampled by an acquisition function which can maximise either "exploration" or "exploitation" which come at a trade-off. Maximising exploitation will cause the algorithm to access the best performing training examples faster but to learn slower. In contrast, maximising explo-

ration will favour data acquisition which helps the model learn faster but might not immediately return the best points in the space. genome-wide screens also present this exploration-exploitation tradeoff. For example, while ALE is powerful at sampling the genomic space to isolate the best performing strains, it is less adapted at generating learnings on the cause of the improvement. On the other hand, SGA has generated some of the most comprehensive biological datasets which are regarded as ground truth by the biological community. However, constraints imposed by the genetic background of the available library (most frequently S288c) and the lack of flexibility in the modifications available might explain why SGA yet has not been used in a strain engineering context.

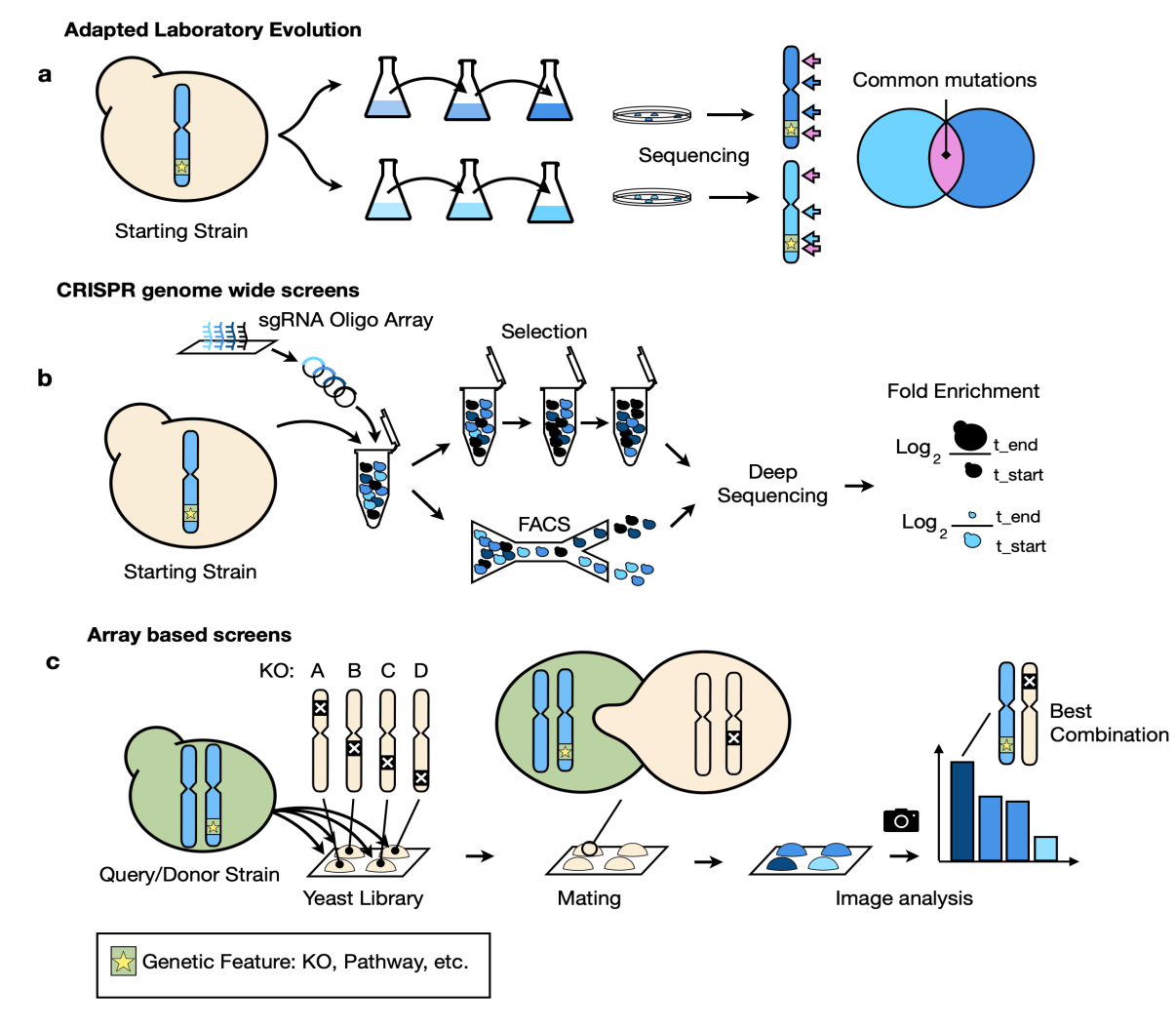


Figure 1.2: **Comparison of three high throughput workflows exploring genome-wide modifications**

a) Adapted laboratory evolution passages microbial populations in challenging growth conditions. The evolved strains are sequenced and convergent mutations among different populations can help edentify the cause of the adaptation. **b)** A typical CRISPR genome-wide screen uses array-based DNA oligo synthesis to build a library of $\approx 10^4$ sgRNAs introducing genome-wide modifications in a starting strain. The best performing strains are selected by a growth challenge or by FACS and their enrichment is revealed by deep sequencing of the sgRNA (or the associated barcode). **c)** Arrayed screens such as SGA and CRI-SPA systematically combine genetic loci in physically separated yeast strains which phenotypes can be analysed by image analysis.

1.5.1 ALE

ALE harnesses the power of evolution to isolate microbial strains with desired traits. Although not a genome editing tool *per se*, ALE efficiently samples the genomic space for modifications and combinations of modifications causing an improvement in a desired phenotype. In its simplest form, a starting strain is passaged in challenging growth conditions for an extended period of time allowing for random mutations to be selected, leading to the emergence of strains better adapted to the conditions (Mans et al. 2018; Shepelin et al. 2018, **fig. 1.2a)**. Because ALE selects better fitness (i.e. faster growth given the specific conditions) it is particularly suited for adapting strains to use alternative substrates or to better tolerate stress (Lennen et al., 2023; Du et al., 2020). Biosensors also offer the possibility to translate the synthesis of a small molecule into a fitness advantage enabling directed evolution of improved titers (Rugbjerg et al., 2018; Xiao et al., 2016; Leavitt et al., 2017; Williams et al., 2016; Raman et al., 2014).

The stochastic nature of the genetic modifications occurring in ALE allows for a free exploration of the host genome. In contrast to the edits possible in semi-targeted screens which are constrained to the size of the oligo, ALE can produce any type of mutation (duplication, substitution, insertion) anywhere along the host genome. ALE is not constrained by human biological knowledge which is still incomplete even for model organisms. This enables ALE to find complex "solutions" in the biological space possibly more efficiently than human design. For example, converting model organisms into chemoautotrophs requires significant rewiring of their metabolism which greatly impacts growth rate. Here, ALE can be used after this first step of rational design to find additional modifications adapting the host to its new metabolism (Gassler et al., 2020; Gleizer et al., 2019). Finally, ALE incrementally combines multiple genetic edits efficiently navigating the exploding combinatorial space. In contrast, combining edits with CRISPR screen requires specific rounds of iterative design (Lian et al., 2017, 2019).

Returning to our active learning analogy, ALE is perhaps the best strategy to "exploit" the genetic landscape, i.e. to find combinations of genetic edits improving strains traits. However, the 'exploration' capacity of ALE, its ability to generate reusable learnings on the mechanisms explaining the improvements, is challenged by the difficulty to disentangle adaptive 'driver' from non-adaptive 'passenger' mutations (Tenaillon et al., 2016; Wang et al., 2020). The general strategy to recognise adaptive drivers is to sequence evolved strain isolates and to identify converging mutations common to multiple isolates (Caspeta et al. 2014; Sandberg et al. 2019, **fig. 1.2a)**. The drop in DNA sequencing cost now supports the sequencing of hundreds of strains (Lennen et al., 2023) which mutations can be catalogued in a specialised database helping in the interpretation of sequencing results (Phaneuf et al., 2019). However, evolved populations tend to accumulate the same mutations (Caspeta et al., 2014). This implies that cheaper sequencing is shifting the bottleneck on the ability of ALE experiments to run hundreds of parallel evolving populations. Finally, ALE experiments typically run for weeks to months (Mavrommati et al., 2022) which rules out their use within a rapidly iterating Design-Built-Test-Learn cycles.

1.5.2 CRISPR-Based Screens

S. cerevisiae DNA repair system boasts an efficient homologous recombination machinery which has facilitated the engineering of its genome before the CRISPR revolution (Jasin and Rothstein, 2013). The introduction of CRISPR in yeast has made this process even more efficient and markerless (DiCarlo et al., 2013). To edit *S. cerevisiae's* genome with CRISPR, a user simply needs to design a sgRNA which guides the nuclease to the restriction site and a repair template which the host uses for homology-directed repair. CRISPR can be used for the integration of multiple DNA fragments at once or the simultaneous edits of multiple sites, i.e. multiplexing (Jakočiūnas, Bonde, Herrgård, Harrison, Kristensen, Pedersen, Jensen and Keasling, 2015; McCarty et al., 2020). State of the art multiplexing now achieves up to 8 simultaneous disruptions with close to 87% efficiency (Zhang et al., 2019). Efficiency, that is the proportion of trans-

formants harbouring all edits, is known to drop with the number of edits or if the sgRNA are not highly efficient (Ryan et al., 2014; Jakočiūnas, Bonde, Herrgård, Harrison, Kristensen, Pedersen, Jensen and Keasling, 2015; Jakociunas, Rajkumar, Zhang, Arsovska, Rodriguez, Jendresen, Skjødt, Nielsen, Borodina, Jensen et al., 2015). Indeed, Zhang et al, observed their efficiency to drop to 23% for 6 simultaneous edits when using untested sgRNAs (Zhang et al., 2019). This must be remembered when benchmarking methods since high throughput CRISPR editing relies on untested sgRNAs

Array-based DNA oligo synthesis has been critical in opening genome-wide engineering with CRISPR. Tens of thousands of programmatically designed sgRNAs can provide genome-wide CRISPR targeting in a single transformation. To enable DNA editing in *S. cerevisiae*, molecular biologists must ensure that a matching repair template is delivered together with the sgRNA to the recipient cell. This can be achieved by fitting both the sgRNA target sequence and a short repair template on the same oligo. The oligo is then cloned on a high copy plasmid which provides both the coding sequence for the sgRNA and the DNA template for repair. For example, the Zhao lab has taken advantage of the self-cleaving mechanism occurring during the maturation of the pre-crRNA into creRNA to place the repair template upstream of the guide on the same oligos. The repair template is co-transcribed with the pre-crRNA but is cleaved during crRNA maturation (Bao et al., 2015, 2018; Lian et al., 2019) A transposon-coupled strategy called CRISPEY provided a creative alternative to provide a DNA repair template in abundance. In CRISPEY, a portion of a retron element was fused to the repair-sgRNA pair to supply the DNA template in the nucleus through reverse transcription (Sharon et al., 2018). In addition, adapters and restriction sites can be added to the design of oligos to support further assembly workflow for adding additional features to the gRNA:template library. One such example is the addition of unique barcodes which can be linked to the repair templates (Sadhu et al., 2018; Roy et al., 2018).

In a brief parenthesis, we note that other mechanisms than CRISPR such as RNA interference and recombineering have been massively successful when coupled to array-based oligo synthesis for genome-wide screening (Si et al., 2015; Wang et al., 2009; Wannier et al., 2021). For example, recombineering harnesses lambda-red bacteriophage's recombination machinery to insert short DNA fragments at the lagging strand of the replication fork of *E. coli* (Ellis et al., 2001). This DNA integration method was successfully coupled to oligo DNA synthesis to mediate simultaneous edits in MAGE (Wang et al., 2009).

Resolving CRISPR based genome-wide screens

In tandem with deep sequencing, CRISPR screens can be performed in one pot enabling the testing of tens of thousands of strains without requiring robotic or microfluidic equipment. Deep sequencing is performed before and after a growth challenge or a cell sorting step and the relative sequence enrichment is used to find deleterious and advantageous edits **(fig. 1.2b)**. CRISPR couples particularly well with deep sequencing because the sgRNA itself can be used as a barcode to quantify enrichments (Garst et al., 2017; Bao et al., 2018; Lian et al., 2019; Bock et al., 2022). However, despite the advance of *in silico* sgRNA prediction tools (Chuai et al., 2017) and because of the oligo synthesis error rate, the perfect efficiency of sgRNAs cannot be assumed. It is therefore common to design multiple guides for the same gene target. An alternative strategy is to barcode the repair template itself during the library preparation (Roy et al., 2018; Sadhu et al., 2018). This links the barcode to its edits more confidently and alleviates the need to maintain the plasmid in the library after editing.

CRISPR screens offer the possibility to both operate on a short sequence in a saturating manner (e.g. a coding sequence) or to introduce mutations genome-wide. For example, saturation mutagenesis of a drug target gene and subsequent exposure to this drug can reveal mutational hotspots frequently granting drug resistance (Roy et al., 2018; Garst et al., 2017). In a genome-wide example, Sadhu et al. introduced premature stop codons along the CDS of more than a thousand yeast essential genes. This revealed that

some of the enriched genes had been wrongfully annotated as essential and others had large dispensable C-terminal regions (Sadhu et al., 2018). In a similar process to ALE, genome-wide screening can also be coupled to growth challenges for identifying mutations favouring resistance to stress. For example, several studies reported genome-wide edits or loss of function mutations promoting thermotolerance or resistance to industrially relevant stressor such as furfural and acetic acid (Garst et al., 2017; Lian et al., 2019; Bao et al., 2018). A notable advantage of CRISPR screens is that multiple edited libraries can be pooled or/and reused for different challenges (Garst et al., 2017).

Genome-wide screens can also be used in a metabolic engineering context to isolate strains with improved product synthesis capacity. To achieve this, product synthesis needs to be directly quantifiable by colorimetry (Barbieri et al., 2017; Wang et al., 2009; Savitskaya et al., 2019; Yang et al., 2018; Özaydın et al., 2013) or translated into a fluorescence signal by a biosensor (Bowman et al., 2021) or an antibody (Lian et al., 2017). This allows the recovery of the best strains by fluorescence activated cell sorting (FACS, or fluorescent activated droplet sorting) or simply by plating (Bowman et al., 2021; Wang et al., 2009). Further, a beneficial modification can be used as a starting point for generating a new library and a new screen, in an incremental process identifying combinations of edits. As mentioned in section 1.4, this directed evolution procedure, pioneered by the Zhao lab, often identifies sets of synergistic mutations and single mutations tested on their own show little to no effects (Lian et al., 2019; Bao et al., 2018).

In summary, CRISPR directed screens have proven capable of both generating learnings and generating better strains. They therefore strike an excellent trade-off between exploration and exploitation. We now turn to SGA, arguably the most powerful exploration genome-wide screen strategy.

1.5.3 Synthetic Gene Array (SGA)

SGA is a high throughput method based on arrayed yeast mating and sporulation to co-segregate genetic features present in two parents into a haploid progeny and to quantify their interactions. Most famously, SGA was used to combine double gene KO by crossing a single KO strain (the query strain) to the strains of Yeast Knock-Out (YKO) library (the array strains) (Kofoed et al., 2015; Tong et al., 2001; Costanzo et al., 2016). A yeast library is an arrayed collection of strains in which Open Reading Frames have been systematically modified, barcoded and marked with a dominant marker. *S. cerevisiae* now boasts numerous libraries, including the YKO library (Giaever et al., 2002), the temperature sensitive conditional KO library (Kofoed et al., 2015), the YETI inducible overexpression library (Arita et al., 2021), and the SWAT libraries enabling a user to rapidly introduce modifications of their own design (Yofe et al., 2016; Weill et al., 2018).

The molecular mechanisms enabling SGA to combine two mutant loci are orchestrated by a sequence of pinning steps on different selective media. SGA starts by crossing a *Mat-α* query strain which mutant locus is marked by the dominant marker *NAT-MX* to the *Mat-a* array strains of the YKO library which mutant loci are marked by the dominant marker *KAN-MX*. The mated diploids are pinned on four different media inducing sporulation, haploid selection and single mutants selection, and double mutants selection (Kuzmin et al. 2014, **fig. 1.3a)**. Haploid selection is made possible by the KO of *CAN1* and *LYP1* in the query strain, which grant recessive resistance to canavanine and thialysine, respectively. In addition, the mutated *CAN1* locus also harbors a cassette in which *Schizosaccharomyces pombe's HIS5* is expressed under the control of the *Mat-a* specific *STE2* promoter. Counter selection on medium containing canavanine and thialysine but lacking histidine is designed to kill diploid cells which still possess a functioning version of any of *CAN1* or *LYP1* and select haploid of *Mat-a* type.

The pinning of colony arrays in SGA is performed with a robot which maintains the physical separa-

tion and positions of all the strains in the YKO **(fig. 1.2c)**. This format can accommodate the testing of the YKO ($\approx$ 4,500 strains) in quadruplicate on thirteen 1536 agar omnitrays. Although this throughput might appear modest in the age of deep sequencing, SGA's format has the unmatched advantage of systematically testing all the strains in a library separately, generating the most comprehensive experimental datasets. From a proof-of-concept screening 8 non-essential genes (Tong et al., 2001), a decade long effort by the Boone and Andrews lab enabled the phenotyping of combinations of millions of double (Costanzo et al., 2016) and hundred thousands of triple KOs of both non-essential and essential genes (Kuzmin et al., 2018). The SGA interaction data is a mine of biological information and is often used as ground truth to benchmark metabolic models (Szappanos et al., 2011; Alzoubi et al., 2019) and train machine learning algorithms (Yu et al., 2016; Ma et al., 2018).

Besides the classical double mutant combination, SGA is a platform technology from which a number of variants have been spun out (Pan et al., 2004, 2006; Snitkin et al., 2008; Jaffe et al., 2017; Mullis et al., 2018). For example, "X-gene" genetic analysis (XGA) is a more recent adaptation of SGA which explores higher orders of combinations. Here, the wild-type query strain is barcoded and mated with the array strain which is knocked-out for several G genes. Meiotic segregation of the G genes yields a pool of progeny cells harbouring random gene combinations drawn from G. Colonies are picked in individual wells and their combinations are associated with the unique barcode via the row-column PCR (RC-PCR) method. Briefly, in RC-PCR the forward and backward primers are barcoded with column and row specific barcodes, respectively. Primers anneal to the same priming sites on the barcoded deleted locus of the YKO library. Each barcoded knock-out is assigned a Row:Column specific pair of barcoded primers, allowing the identification of its position on a plate with amplicon sequencing (Yachie et al., 2016). Once the various combinations in the library are linked to a unique barcode, the library is pooled and grown in challenging conditions and the enrichment of each barcode is used as a fitness measure (Celaj et al., 2020).

Resolving SGA screens.

The SGA workflow outputs 1536 arrays of physically separated colonies of yeast double mutants **(fig. 1.2c)**. In theory, if an interaction between the query and array mutant loci causes a change in fitness, the difference in colony size should be easily detectable by image analysis. However, a number of screen artefacts other than the genomic background of the strains might affect colony growth. These artefacts need to be corrected to make every double mutant comparable with each other.

High density screen artefacts such as plate-to-plate variability, neighbour effects and edge effects are well described and can effectively be removed (Baryshnikova et al., 2010; Wagih et al., 2013). Here, because genetic interactions are rare, it can be assumed that a large enough set of colonies must have a median fitness equal to that of the WT. Therefore, if the median of a same set of colonies (e.g. colonies on the same plate) departs from the screen median, the difference can be assumed to be caused by a common artefact and removed (Baryshnikova et al., 2010; Collins et al., 2006). This assumption is behind all the corrective functions developed by the Boone and Andrews lab. The specificity of the correction arises from the set of colonies which are compared. For example, to correct differences in agar quality, a convolution is applied to measure the median of a local 7by7 patch of colonies. The scores of all the colonies in this subset are rescaled so that the median of the 7by7 patch matches the median of the plate to remove local agar gradients (Wagih et al., 2013).

Last, since a genetic interaction of a double KO is defined as a fitness which differs from an expectation based on the fitnesses of the single KOs, defining the expectation requires quantifying the fitness of the single mutants. Once again the "median assumption", can be employed to infer the fitness of single mutants from double mutants arrays. Since most gene combinations are neutral, the fitness for single

mutant A can be assumed to be the median fitness of all pairs of genes including A (A:B, A:C ... A:Z) (Baryshnikova et al., 2010; Collins et al., 2006). SGA studies can then define the expected fitness as the product of the single mutants fitness. The interaction score then becomes ϵ = fitness(AB) - fitness(A) x fitness(B). Besides this quantification scheme, the S-score is also a popular alternative to quantify interactions (Kumar et al., 2016; Schuldiner et al., 2005; Butland et al., 2008; Guénolé et al., 2013; Martin et al., 2015). The S-score formula is similar to that of the t-value. It accounts for data variability and computes the confidence with which an interaction can be called 'epistatic'.

In summary, the SGA screen and its accompanying data analysis constitute a highly optimised platform enabling the systematic testing of pairs of interacting loci. SGA generates the most complete datasets making it a popular exploration tool for fundamental genetic studies. We are not aware of its use as an exploitation tool to generate strains with improved traits. One possible reason is that the genetic background in which interactions are tested is constrained by the library (usually S288c).

To summarise on genetic screens, ALE, CRISPR-based screens and SGA strike different exploration-exploitation abilities. At one end of the spectrum, ALE is particularly powerful at exploitation, i.e. generating improved strains. CRISPR based screens hit a balanced exploration-exploitation trade off. They can both be coupled to selection methods to isolate strains with better performances and with deep-sequencing to inform on the types of edits causing improvements. Last, SGA systematically tests all variants in a library generating complete datasets and represents therefore the best exploration alternative. This exploration ability is particularly suited for the mapping of the genetic interaction space. Indeed, the bias for fitness introduced by library pooling in CRISPR-based screen means that more productive edits might be lost if they are accompanied by a fitness defect. Second, the best performing strains might saturate the top limiting the number of beneficial edits which can be recovered (Savitskaya et al., 2019). Being interested in testing host:pathway interactions across all host genes, an array based method such as SGA is therefore particularly adapted.

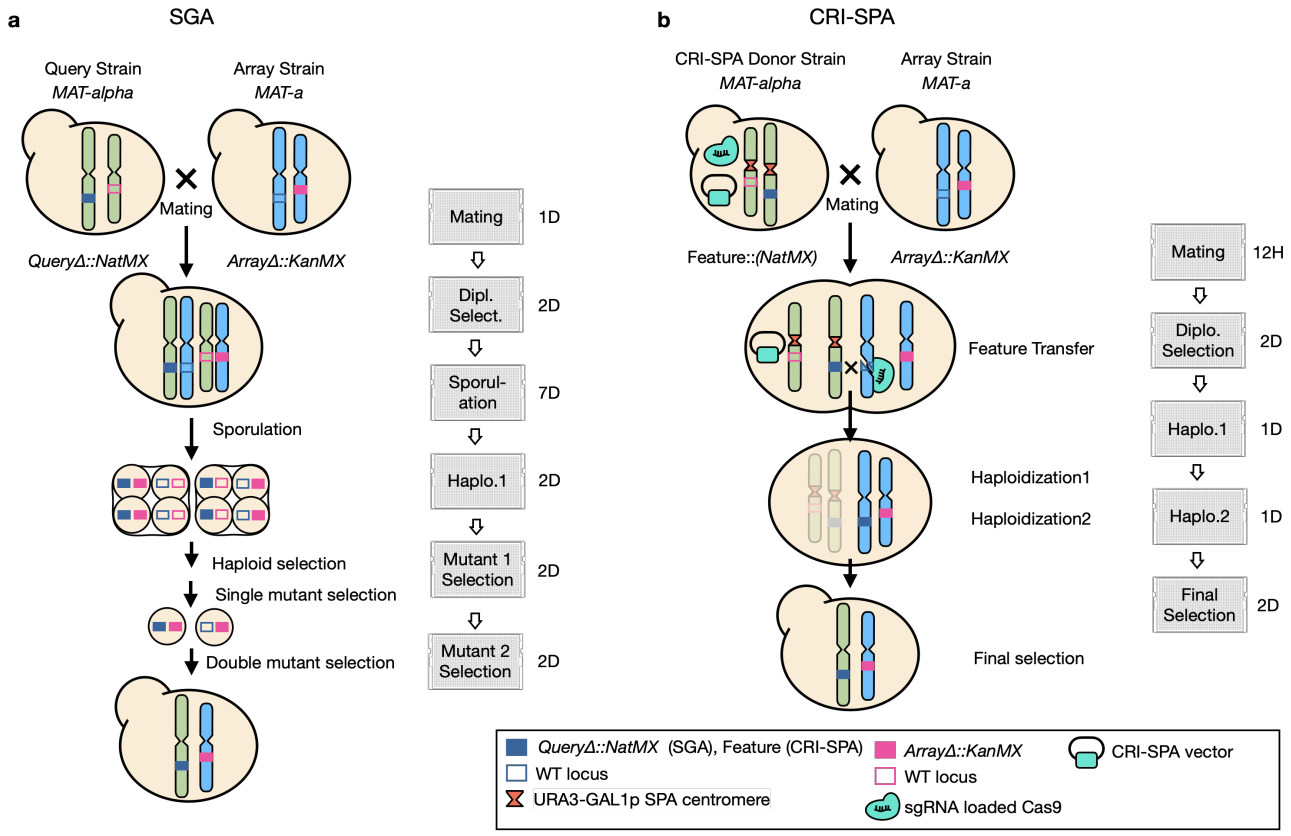


Figure 1.3: **Overview of the SGA and CRI-SPA procedures.**
SGA and CRI-SPA both combine a locus (SGA: genrally a KO, CRI-SPA: a genetic feature) to a modified locus present in the strains of a library through yeast mating. **a)** In SGA, co-segregation of the double mutants is driven by meiosis and multiple rounds of selection for haploids, single, and double mutants. The procedure requires 6 pinning steps and 16 days of incubation.(Adapted from Kuzmin et al. 2014). **b)** In CRI-SPA, the transfer of the genetic feature is mediated by a CRISPR-Cas9-induced cut of the recipient site, resulting in its repair using the donor chromosome as a repair template. The chromosomes of the donor strain are then *specifically ablated* by activation and counter selection of their SPA centromeres. The procedure requires 5 pinning steps and 6 to 7 days incubation.

1.5.4 CRI-SPA

Planning on using SGA at the beginning of my PhD, I coincidentally heard about a similar method, CRI-SPA, in development in the Mortensen Group at DTU Bioengineering (Olsson, 2020; Coumou, 2016). The format of CRI-SPA is similar to that of SGA as it systematically introduces a genetic feature to a yeast library in an array format. However, the mechanisms driving the delivery of the feature in the library strains does not rely on meiosis, granting CRI-SPA significant advantages.

Like SGA, CRI-SPA starts by mating the CRI-SPA donor strain to the strains of a yeast library. The donor strain harbours a CRI-SPA vector and the mated diploids are select by selecting for the markers harboured by the vector and the modified locus in the library background. In the diploid state, **CRI**SPR-Cas9 encoded by the donor strain is guided by the sgRNA encoded by the CRI-SPA vector to the recipient site where it creates a double strand break (DBS). The DBS is repaired by the host homologous recombination machinery using the donor chromosome as a template. Importantly, the genetic feature is located on the same site on the donor chromosome but does not harbour the sgRNA recognition site **(fig. 1.3b)**. Hence, homologous repair leads to a gene conversion event transferring the genetic feature into the recipient site which is no longer targeted by CRISPR.

After feature transfer, the chromosomes of the donor strain are selectively destabilised and deleted from the diploid allowing the progeny to return to a haploid state without meiosis. This mechanism relies of **S**eleclive **P**loidy **A**blation (SPA) enbaled by conditional centromeres present on all chromosomes of the donor strain (Reid et al., 2011). Each conditional centromere is equipped with the inducible galactose promoter *GAL1* which induction interferes with the centromere's mitotic segregation function (Chlebowicz-Śledziewska and Śledziewski, 1985). Centromere are also equipped with the *URA3* counter selectable marker. *GAL1p* activation and *URA3* counter selection leads to the selective loss of chromosomes harboring the SPA centromere (Reid et al., 2008).

The ability of CRI-SPA to return to the haploid state without meiosis confers the method a number of advantages over SGA. Most importantly, genetic exchange between homologous chromosomes during meiosis through crossovers and gene conversions means that the genetic material of the haploids generated by SGA will be a mixture of both parents chromosomes (Mancera et al., 2008). The query strain must therefore have the same genetic background as the library (S288c) to ensure that the interaction between the combined loci is not masked by the uncontrolled combination of other heterologous genetic elements (Kuzmin et al., 2014). Further, co-segregation of mutants which loci lie on the same chromosome necessitates a crossover between them. A crossover between neighbouring loci being unlikely, some gene pairs are impossible to combine by SGA (Jorgensen et al., 2002). Practically, avoiding meiosis also allows CRI-SPA to be more than twice a fast as SGA. Indeed, meiosis, the longest step in SGA, takes seven days which is the time taken for the entire CRI-SPA procedure **(fig. 1.3)**.

A final advantage of CRI-SPA is that the CRISPR-mediated DBS might itself serve as a counter selection for strains which did not integrate the genetic feature. This counter selection alleviates the need to mark the feature with a selectable marker opening the possibility to perform multiple rounds of edits or multiplex editing with CRI-SPA.

Given the many advantages of CRI-SPA over SGA, I selected this platform to support the screening of genetic interactions at high throughput. I started a collaboration with Pr. Mortensen to continue the development of CRI-SPA as part of my own project. The description of CRI-SPA and its use as an exploration tool in metabolic engineering are described in chapter 3 the first paper of this thesis.

1.6 Betaxanthins and *cis-cis*-muconic acid. Two metabolic engineering case studies.

1.6.1 Linking production to a phenotype

High throughput genome editing methods now produce thousand to millions of strains variants and their phenotyping has become the new bottleneck for metabolic engineering. Analytical chemistry methods such as Laser Assisted Ambient Ionization MS (LA-REIMS) are reaching higher throughput allowing for the processing of one to ten samples in a minute (Fu et al., 2022; Plekhova et al., 2021). For example, Gowers et al. were able to screen for improved production of betulinic acid at $\approx$ two colonies per minute with LA-REIMS (Gowers et al., 2019). This pace is still not sufficient for alleviating the phenotyping bottleneck and the false positive rate reported (75%) was high.

Ideally, differences in synthesis ability caused by genetic edits would be directly quantifiable through a conspicuous phenotype associated with each strain variant. As mentioned above (section 1.5.2), this is feasible for a handful of products which are fluorescent or colorful. However, synthesis of these compounds is already highly optimised as they are often employed as case studies in the development of new high throughput screens. Alternatively, the availability of a wide variety of biosensors now offers to bridge the

genotype-to-phenotype bottleneck for otherwise inconspicuous products (Koch et al., 2019; Kaczmarek and Prather, 2021). Advantageously, biosensors scale to any throughput at no additional cost and their readout is compatible with existing lab equipment (e.g. fluorescent imagers, plate readers, FACS).

In this thesis, I have applied the screening of genetic interactions to two metabolic engineering targets. When I took-over the lead on CRI-SPA, it had been developed to screen for the production of betaxanthins, a family of yellow plant pigments. Productivity of betaxanthins could be extracted directly from screen plates by measuring colony yellow intensity with image analysis. In a second version, I extended the applicability of CRI-SPA by coupling the synthesis of the colorless small molecule, *cis-cis*-muconic acid, to a fluorescent signal thanks to an inducible biosensor. In the last part of this introduction, I briefly introduce the metabolic engineering context surrounding the synthesis of these two compounds in *S. cerevisiae.*

1.6.2 Betaxanthins

Betaxanthins are a subgroup of betalains, a large family of plant pigments responsible for the red, violet, pink and yellow colors observed in the flowering plant species of the order Caryophyllales. Betaxanthins are synthesised from tyrosine which is hydroxylated into 3,4-dihydroxy-L-phenyalanine (L-DOPA). The aromatic ring of L-DOPA is opened by an extradiol cleavage which is followed by a spontaneous recyclization leading to the formation of betalamic acid (Sasaki et al., 2009). Finally, the unstable aldehyde group of betalamic acid spontaneously binds with an amine group of an available amino acid resulting in the formation of betaxanthins **(fig. 1.4)**. Because this last reaction occurs none specifically, betalamic acid can link to various amino acids. Hence, Betaxanthins are a family of compounds rather than a specific molecule (Polturak and Aharoni, 2018) but for simplicity we will henceforth refer to this diversity as 'bethaxantin'.

Figure 1.4: **Heterologous Synthesis of Betaxanthins in Yeast**
Yellow plant pigment betaxanthins can be synthesised from L-tyrosine by expression of CYP76AD1 from *Beta vulgaris* and DOPA dioxygenase (DOD) from *Mirabilis jalapa's*. The last reaction between betalamic acid and available amine groups is spontaneous.

The yellow color and fluorescence of betaxanthin was first exploited in yeast by the Dueber Lab as an enzymatic biosensor to identify an enzyme able to convert L-tyrosine into L-DOPA. (DeLoache et al., 2015). They first expressed *Mirabilis jalapa's* DOPA dioxygenase (DOD) converting L-DOPA into betaxanthin which could be detected with flow cytometry. This enabled the authors to rapidly test enzymes synthesising L-DOPA from L-tyrosine and found that the cytochrome P450 CYP76AD1 from the sugar beet *Beta vulgaris* was able to catalyse this step. They further improved CYP76AD1 activity by selecting error prone PCR-generated variants by simple visual selection of colonies growing on Petri dishes.

Increasing the synthesis of betaxantins and its precursors in yeast is relevant for a number of metabolic engineering applications. First, L-DOPA is itself used as treatment to slow the progress of Parkinson's disease (Rai et al., 2021). In addition, L-DOPA is a precursor of benzylisoquinoline alkaloids, a family of molecules with potent analagesic properties including opioids. Abundant supply of L-DOPA is of interest as a precursor feeding into the synthesis of opioids which full pathway was introduced in yeast by

the Smolke Group (Galanie et al., 2015). Further, host modifications improving betaxanthin synthesis may do so by improving the activity of the cytochrome P450 CYP76AD1 (Savitskaya et al., 2019; Jiang, Dong, Liu, Shi, Wang, Tao, Liang and Lian, 2021), a class of enzyme with broad industrial relevance (Jiang, Huang, Cai, Xu and Lian, 2021). Finally, betaxanthin itself has attracted growing attention for its antioxidant and coloring properties as a food supplement (Polturak and Aharoni, 2018).

In summary, betaxanthin produce a strong yellow phenotype in yeast and its synthesis requires the expression of only two heterologous coding sequences. Given its broad industrial relevance, it represents a great test bed for the development of CRI-SPA.

1.6.3 *cis-cis*-Muconic Acid

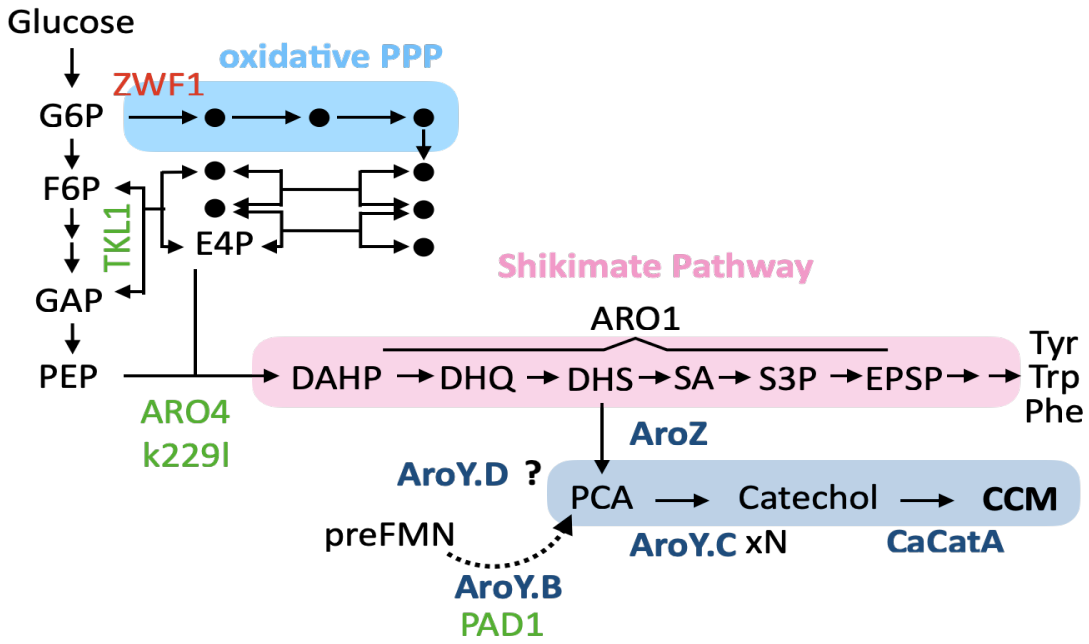


Figure 1.5: **Heterologous Synthesis of CCM in Yeast**
CCM can be synthesised in yeast from the shikimate pathway intermediate DHS by the set of heretologous enzymes in bold blue. Strategies improving metabolic flux towards CCM include TKL1 overexpression and ZWF1 KO, expression of the feed back resistant $ARO4^{k229l}$ and decoupling of some of the pentafunctional ARO1 domains. The decarboxylation of PCA to catechol is a bottleneck which can be alleviated by overexpression of PAD1 from S288C and multiple integration of AROY-BC. G6P: glucose-6-phosphate, F6P: fructose-6-phosphate, GAP:glyceraldehyde-3-phosphate, PEP: phosphoenolpyruvate, E4P: erythrose 4-phosphate, DAHP: 3-Deoxy-D-arabinoheptulosonate 7-phosphate, DHQ:3-Dehydroquinic acid, DHS: 3-dehydroshikimate, SA: shikimic acid, S3P: SA-3-phosphate, EPSP: 5-enolpyruvylshikimate-3-phosphate, PCA: protocatechuic acid.

Cis-cis-muconic (CCM) acid is an important industrial platform chemical from which a variety of value added derivatives can be manufactured, including the nylon precursor adipic acid (Khalil et al., 2020; Xie et al., 2014). A number of microbial hosts and enzymatic routes have been used for the *de novo* synthesis of CCM, all using aromatic precursors synthesised by the shikimate pathway which natural function is to provide the aromatic amino acids tyrosine, phenilalanine and tryptophan (Thompson et al., 2018). In yeast, CCM was synthesised from the shikimate pathway intermediate 3-dehydroshikimate through three enzymatic steps (Curran et al., 2013*a*; Weber et al., 2012).

In the past decade, numerous studies have sought to increase CCM titers in yeast. A possible strategy is to improve metabolic flux towards the shikimate pathway and into the heterologous CCM branch.

For example, increasing the availability of the precursors phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P) can be achieved by KO of *ZWF1* and overexpression of *TKL1*, favouring metabolic entry into glycolysis rather than the oxidative pentose phosphate pathway (Curran et al., 2013*b*). In addition, expression of the aromatic amino acid feedback resistant $ARO4^{k229l}$ increases flux into the pathway regardless of aromatic amino acid levels (Curran et al., 2013*b*). Next, the increase in shikimate pathway flux can be committed into the heterologous CCM branch by reducing the efficiency of the downstream reactions which compete for DHS with CCM synthesis **(fig. 1.5)**. Since the pentafunctional Aro1 catalyses the five consecutive steps from DHAP to EPSP **(fig. 1.5)**, the enzymatic domains catalysing DHQ and DHS synthesis need to stay functional to support CCM. Moreover, the full KO of the steps downstream of DHS would result in aromatic amino acid auxotrophy which is undesirable in an industrial context. Conveniently, the DHS dehydrogenase function is encoded on the C-terminal part of *ARO1*. Taking advantage of this feature, Leavitt et al. truncated the C-terminal and reintroduced *E. coli*'s DHS dehydrogenase, AroE. This enabled the authors to overexpress the truncated Aro1 to drive high flux towards DHS while keeping lower levels of AroE expression to reduce the depletion of DHS for aromatic amino acid synthesis (Leavitt et al., 2017). A similar decoupling strategy was also implemented by Pyne et al. who reduced Aro1 levels with a degradation tag while increasing DAHP to DHS flux by over expressing *E. coli*'s *AroB* and *AroD*

Besides favouring precursor fluxes towards the CCM pathway, the decarboxylation of PCA into catechol is a key pathway target as it has been repeatedly reported to be rate limiting (Curran et al., 2013*b*; Horwitz et al., 2015; Weber et al., 2017; Leavitt et al., 2017). While the *Klebsiella pneumoniae*'s AroY-BCD gene clusters was originally introduced in *S. cerevisiae* to catalyse this step, the decarboxylation is only catalysed by AroY-C. AroY-B catalyses the synthesis of the cofactor prenylated flavin mononucleotide (FMN) which is necessary for AroY-C activity (Payer et al. 2017, **fig. 1.5)**. However, the S288c endogenous gene *PAD1* can also catalyse this step and AroY-B is dispensable in this background (Weber et al., 2017). The role of AroY-D is still unknown and it was found to be actually detrimental for catechol synthesis in a CEN.PK background (Weber et al., 2012). A common strategy to relax the decarboxylation bottleneck is to overexpress the decarboxylase by multicopy integration of *AROY-C* into Ty4 sites (Skjoedt, Snoek, Kildegaard, Arsovska, Eichenberger, Goedecke, Rajkumar, Zhang, Kristensen, Lehka et al., 2016; Snoek et al., 2018) or by overexpression of *PAD1* (Pyne et al., 2018). More recently, Jensen et al. isolated a more productive variants of AroY-B and Aroy-C by evolving the two enzymes on orthorep (Ravikumar et al., 2018) and sorting the best producers by FACS. This strategy relied on a CCM biosensor to convert CCM level into a a fluorescence signal supporting the selection of the best variants with FACS (Jensen et al., 2021).

The CCM biosensor, an important tool for CCM research in yeast, is the result of two engineering efforts from our group. This biosensor uses BenM, a LysR-type transcriptional regulator naturally responsive to CCM activating *benABCDE*, the benzoate degradation operon in *Acinetobacter* spp. Transcriptional activation of the biosensor relies on a BenM responsive promoter created by inserting the BenM operator BenO in *S. cerevisiae*'s CYC1 promoter. In its first description, Skjoedt et al. identified the best insertion location of BenO along *CYC1p* and tuned the expression level of BenM, resulting in CCM activation with lowest expression level and the highest dynamic range. Its dynamic range was then improved by error-prone PCR of the CCM-binding domain of BenM followed by a toggled FACS selection. Toggled FACS, performs two rounds of selection for cells showing the lowest background and the highest activation in the absence and presence of CCM, respectively. This enabled the authors to isolate a biosensor with a 4 fold dynamic range when activated with 1.4 mM (200 mg/L) CCM (Skjoedt, Snoek, Kildegaard, Arsovska, Eichenberger, Goedecke, Rajkumar, Zhang, Kristensen, Lehka and others, 2016). A subsequent study, also employed toggled FACS to further increase the dynamic (fluorescence output) and operational (CCM concentrations input) ranges of the biosensor. This resulted in an updated biosensor version with a close to fifteen-fold activation range and an operational range of 0.01 to 5 mM CCM (Snoek et al., 2020).

The extensive engineering of a reliable BenM biosensor has been instrumental to a number of screens seeking to improve CCM titers in yeast. Mutltiple studies, used it to select more productive strains by FACS (Wang et al., 2020, 2021; Jensen et al., 2021) or by coupling CCM synthesis to the expression of an antibiotic resistance for fitness-based selection (Snoek et al., 2020).

In summary, CCM biosynthesis in yeast is well described and has been the subject of multiple metabolic engineering efforts over the past 10 years. Yet, the strain engineering rationals for further improving its yield seem to have been exhausted. In addition, CCM boasts a highly optimised biosensor which reliability is supported by its repeated use by several groups. Taken together, CCM biosynthesis represents an excellent case-study for testing the potential of leveraging host:pathway interactions for improving product titers in yeast. In the second manuscript of this thesis, we describe the adaptation of CRI-SPA to the screening of CCM biosynthesis thanks to the BenM biosensor.

Bibliography

Alzoubi, D., Desouki, A. A. and Lercher, M. J. (2019), 'Flux balance analysis with or without molecular crowding fails to predict two thirds of experimentally observed epistasis in yeast', *Scientific reports* **9**(1), 1–9.

Arita, Y., Kim, G., Li, Z., Friesen, H., Turco, G., Wang, R. Y., Climie, D., Usaj, M., Hotz, M., Stoops, E. H. et al. (2021), 'A genome-scale yeast library with inducible expression of individual genes', *Molecular Systems Biology* **17**(6), e10207.

Bakerlee, C. W., Nguyen Ba, A. N., Shulgina, Y., Rojas Echenique, J. I. and Desai, M. M. (2022), 'Idiosyncratic epistasis leads to global fitness–correlated trends', *Science* **376**(6593), 630–635.

Bank, C. (2022), 'Epistasis and adaptation on fitness landscapes', *Annual review of ecology, evolution, and systematics* **53**, 457–479.

Bank, C., Matuszewski, S., Hietpas, R. T. and Jensen, J. D. (2016), 'On the (un) predictability of a large intragenic fitness landscape', *Proceedings of the National Academy of Sciences* **113**(49), 14085–14090.

Bao, Z., HamediRad, M., Xue, P., Xiao, H., Tasan, I., Chao, R., Liang, J. and Zhao, H. (2018), 'Genome-scale engineering of saccharomyces cerevisiae with single-nucleotide precision', *Nature biotechnology* **36**(6), 505–508.

Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., Si, T. and Zhao, H. (2015), 'Homology-integrated crispr–cas (hi-crispr) system for one-step multigene disruption in saccharomyces cerevisiae', *ACS synthetic biology* **4**(5), 585–594.

Barajas, C., Huang, H.-H., Gibson, J., Sandoval, L. and Del Vecchio, D. (2022), 'Feedforward growth rate control mitigates gene activation burden', *Nature Communications* **13**(1), 7054.

Barbieri, E. M., Muir, P., Akhuetie-Oni, B. O., Yellman, C. M. and Isaacs, F. J. (2017), 'Precise editing at dna replication forks enables multiplex genome engineering in eukaryotes', *Cell* **171**(6), 1453–1467.

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S. and others (2010), 'Quantitative analysis of fitness and genetic interactions in yeast on a genome scale', *Nature methods* **7**(12), 1017–1024. Publisher: Nature Publishing Group US New York.

Bateson, W. (1909), 'Mendel's principles of heredity: Cambridge university press', *März 1909; 2nd Impr* **3**, 1913.

Bergelson, J., Kreitman, M., Petrov, D. A., Sanchez, A. and Tikhonov, M. (2021), 'Functional biology in its natural context: A search for emergent simplicity', *Elife* **10**, e67646.

Bock, C., Datlinger, P., Chardon, F., Coelho, M. A., Dong, M. B., Lawson, K. A., Lu, T., Maroc, L., Norman, T. M., Song, B. and others (2022), 'High-content CRISPR screening', *Nature Reviews Methods Primers* **2**(1), 8. Publisher: Nature Publishing Group UK London.

Bonnet, J., Subsoontorn, P. and Endy, D. (2012), 'Rewritable digital data storage in live cells via engineered control of recombination directionality', *Proceedings of the National Academy of Sciences* **109**(23), 8884–8889.

Boo, A., Ellis, T. and Stan, G.-B. (2019), 'Host-aware synthetic biology', *Current Opinion in Systems Biology* **14**, 66–72. Publisher: Elsevier.

Bowman, E. K., Wagner, J. M., Yuan, S.-F., Deaner, M., Palmer, C. M., D'Oelsnitz, S., Cordova, L., Li, X., Craig, F. F. and Alper, H. S. (2021), 'Sorting for secreted molecule production using a biosensor-in-microdroplet approach', *Proceedings of the National Academy of Sciences* **118**(36), e2106818118.

Butland, G., Babu, M., Díaz-Mejía, J. J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A. G., Pogoutse, O. et al. (2008), 'esga: E. coli synthetic genetic array analysis', *Nature methods* **5**(9), 789.

Canton, B., Labno, A. and Endy, D. (2008), 'Refinement and standardization of synthetic biological parts and devices', *Nature biotechnology* **26**(7), 787–793.

Cardinale, S. and Arkin, A. P. (2012), 'Contextualizing context for synthetic biology–identifying causes of failure of synthetic biological systems', *Biotechnology journal* **7**(7), 856–866. Publisher: Wiley Online Library.

Caspeta, L., Chen, Y., Ghiaci, P., Feizi, A., Buskov, S., Hallström, B. M., Petranovic, D. and Nielsen, J. (2014), 'Altered sterol composition renders yeast thermotolerant', *Science* **346**(6205), 75–78.

Celaj, A., Gebbia, M., Musa, L., Cote, A. G., Snider, J., Wong, V., Ko, M., Fong, T., Bansal, P., Mellor, J. C. et al. (2020), 'Highly combinatorial genetic interaction analysis reveals a multi-drug transporter influence network', *Cell Systems* **10**(1), 25–38.

Ceroni, F., Boo, A., Furini, S., Gorochowski, T. E., Borkowski, O., Ladak, Y. N., Awan, A. R., Gilbert, C., Stan, G.-B. and Ellis, T. (2018), 'Burden-driven feedback control of gene expression', *Nature methods* **15**(5), 387–393.

Chan, D. T. C., Baldwin, G. S. and Bernstein, H. C. (2023), 'Revealing the chassis-effect on a broad-host-range genetic switch and its concordance with interspecies bacterial physiologies', *bioRxiv* pp. 2023–02. Publisher: Cold Spring Harbor Laboratory.

Chlebowicz-Śledziewska, E. and Śledziewski, A. Z. (1985), 'Construction of multicopy yeast plasmids with regulated centromere function', *Gene* **39**(1), 25–31.

Chuai, G.-h., Wang, Q.-L. and Liu, Q. (2017), 'In silico meets in vivo: towards computational crispr-based sgrna design', *Trends in biotechnology* **35**(1), 12–21.

Collins, S. R., Schuldiner, M., Krogan, N. J. and Weissman, J. S. (2006), 'A strategy for extracting and analyzing large-scale quantitative epistatic interaction data', *Genome biology* **7**(7), R63.

Costanzo, M., Kuzmin, E., van Leeuwen, J., Mair, B., Moffat, J., Boone, C. and Andrews, B. (2019), 'Global genetic networks and the genotype-to-phenotype relationship', *Cell* **177**(1), 85–100.

Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S. D. et al. (2016), 'A global genetic interaction network maps a wiring diagram of cellular function', *Science* **353**(6306), aaf1420.

Costello, A. and Badran, A. H. (2021), 'Synthetic biological circuits within an orthogonal central dogma', *Trends in biotechnology* **39**(1), 59–71.

Coumou, H. (2016), 'Engineering baker's yeast for the production of aromatic plant-derived compounds', *PhD Thesis, DTU* .

Curran, K. A., Leavitt, J. M., Karim, A. S. and Alper, H. S. (2013*a*), 'Metabolic engineering of muconic acid production in saccharomyces cerevisiae', *Metabolic engineering* **15**, 55–66.

Curran, K. A., Leavitt, J. M., Karim, A. S. and Alper, H. S. (2013*b*), 'Metabolic engineering of muconic acid production in Saccharomyces cerevisiae', *Metabolic engineering* **15**, 55–66. Publisher: Elsevier.

Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., Petzold, C. J., Mukhopadhyay, A., Lee, T. S., Adams, P. D. and others (2013), 'Engineering dynamic pathway regulation using stress-response promoters', *Nature biotechnology* **31**(11), 1039–1046. Publisher: Nature Publishing Group UK London.

Darlington, A. P., Kim, J., Jiménez, J. I. and Bates, D. G. (2018), 'Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes', *Nature communications* **9**(1), 695.

de Lorenzo, V. (2011), 'Beware of metaphors: chasses and orthogonality in synthetic biology', *Bioengineered bugs* **2**(1), 3–7. Publisher: Taylor & Francis.

DeLoache, W. C., Russ, Z. N., Narcross, L., Gonzales, A. M., Martin, V. J. and Dueber, J. E. (2015), 'An enzyme-coupled biosensor enables (s)-reticuline production in yeast from glucose', *Nature chemical biology* **11**(7), 465–471.

DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J. and Church, G. M. (2013), 'Genome engineering in saccharomyces cerevisiae using crispr-cas systems', *Nucleic acids research* **41**(7), 4336–4343.

Domingo, J., Baeza-Centurion, P. and Lehner, B. (2019), 'The causes and consequences of genetic interactions (epistasis)', *Annual review of genomics and human genetics* **20**, 433–460.

Dowell, R. D., Ryan, O., Jansen, A., Cheung, D., Agarwala, S., Danford, T., Bernstein, D. A., Rolfe, P. A., Heisler, L. E., Chin, B. et al. (2010), 'Genotype to phenotype: a complex problem', *Science* **328**(5977), 469–469.

Du, B., Olson, C. A., Sastry, A. V., Fang, X., Phaneuf, P. V., Chen, K., Wu, M., Szubin, R., Xu, S., Gao, Y. et al. (2020), 'Adaptive laboratory evolution of escherichia coli under acid stress', *Microbiology* **166**(2), 141.

Ellis, H. M., Yu, D., DiTizio, T. and Court, D. L. (2001), 'High efficiency mutagenesis, repair, and engineering of chromosomal dna using single-stranded oligonucleotides', *Proceedings of the National Academy of Sciences* **98**(12), 6742–6746.

Fang, G., Wang, W., Paunic, V., Heydari, H., Costanzo, M., Liu, X., Liu, X., VanderSluis, B., Oately, B., Steinbach, M. et al. (2019), 'Discovering genetic interactions bridging pathways in genome-wide association studies', *Nature communications* **10**(1), 4274.

Fernandez-Rodriguez, J., Moser, F., Song, M. and Voigt, C. A. (2017), 'Engineering rgb color vision into escherichia coli', *Nature chemical biology* **13**(7), 706–708.

Friedland, A. E., Lu, T. K., Wang, X., Shi, D., Church, G. and Collins, J. J. (2009), 'Synthetic gene networks that count', *science* **324**(5931), 1199–1202.

Fu, L., Guo, E., Zhang, J., Li, K., Chen, Y. and Si, T. (2022), 'Towards one sample per second for mass spectrometric screening of engineered microbial strains', *Current Opinion in Biotechnology* **76**, 102725.

Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M. and Smolke, C. D. (2015), 'Complete biosynthesis of opioids in yeast', *Science* **349**(6252), 1095–1100. Publisher: American Association for the Advancement of Science.

Garst, A. D., Bassalo, M. C., Pines, G., Lynch, S. A., Halweg-Edwards, A. L., Liu, R., Liang, L., Wang, Z., Zeitoun, R., Alexander, W. G. et al. (2017), 'Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering', *Nature biotechnology* **35**(1), 48.

Gassler, T., Sauer, M., Gasser, B., Egermeier, M., Troyer, C., Causon, T., Hann, S., Mattanovich, D. and Steiger, M. G. (2020), 'The industrial yeast pichia pastoris is converted from a heterotroph into an autotroph capable of growth on co2', *Nature Biotechnology* **38**(2), 210–216.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B. et al. (2002), 'Functional profiling of the saccharomyces cerevisiae genome', *nature* **418**(6896), 387–391.

Gleizer, S., Ben-Nissan, R., Bar-On, Y. M., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A. et al. (2019), 'Conversion of escherichia coli to generate all biomass carbon from co2', *Cell* **179**(6), 1255–1263.

Gowers, G.-O. F., Cameron, S. J., Perdones-Montero, A., Bell, D., Chee, S. M., Kern, M., Tew, D., Ellis, T. and Takáts, Z. (2019), 'Off-colony screening of biosynthetic libraries by rapid laser-enabled mass spectrometry', *ACS synthetic biology* **8**(11), 2566–2575.

Gross, G., Waks, T. and Eshhar, Z. (1989), 'Expression of immunoglobulin-t-cell receptor chimeric molecules as functional receptors with antibody-type specificity.', *Proceedings of the National Academy of Sciences* **86**(24), 10024–10028.

Guénolé, A., Srivas, R., Vreeken, K., Wang, Z. Z., Wang, S., Krogan, N. J., Ideker, T. and van Attikum, H. (2013), 'Dissection of dna damage responses using multiconditional genetic interaction maps', *Molecular cell* **49**(2), 346–358.

Gyorgy, A., Jiménez, J. I., Yazbek, J., Huang, H.-H., Chung, H., Weiss, R. and Del Vecchio, D. (2015), 'Isocost lines describe the cellular economy of genetic circuits', *Biophysical journal* **109**(3), 639–646. Publisher: Elsevier.

Harrison, R., Papp, B., Pál, C., Oliver, S. G. and Delneri, D. (2007), 'Plasticity of genetic interactions in metabolic networks of yeast', *Proceedings of the National Academy of Sciences* **104**(7), 2307–2312.

Horwitz, A. A., Walter, J. M., Schubert, M. G., Kung, S. H., Hawkins, K., Platt, D. M., Hernday, A. D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, S. S. and others (2015), 'Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas', *Cell systems* **1**(1), 88–96. Publisher: Elsevier.

Hou, J., Tan, G., Fink, G. R., Andrews, B. J. and Boone, C. (2019), 'Complex modifier landscape underlying genetic background effects', *Proceedings of the National Academy of Sciences* **116**(11), 5045–5054.

Jacob, F. (1977), 'Evolution and tinkering', *Science* **196**(4295), 1161–1166.

Jaffe, M., Sherlock, G. and Levy, S. F. (2017), 'iseq: a new double-barcode method for detecting dynamic genetic interactions in yeast', *G3: Genes, Genomes, Genetics* **7**(1), 143–153.

Jakočiūnas, T., Bonde, I., Herrgård, M., Harrison, S. J., Kristensen, M., Pedersen, L. E., Jensen, M. K. and Keasling, J. D. (2015), 'Multiplex metabolic pathway engineering using crispr/cas9 in saccharomyces cerevisiae', *Metabolic engineering* **28**, 213–222.

Jakociunas, T., Rajkumar, A. S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C. B., Skjødt, M. L., Nielsen, A. T., Borodina, I., Jensen, M. K. et al. (2015), 'Casemblr: Cas9-facilitated multiloci genomic integration of in vivo assembled dna parts in saccharomyces cerevisiae', *ACS synthetic biology* **4**(11), 1226–1234.

Jasin, M. and Rothstein, R. (2013), 'Repair of strand breaks by homologous recombination', *Cold Spring Harbor perspectives in biology* **5**(11), a012740.

Jensen, E. D., Ambri, F., Bendtsen, M. B., Javanpour, A. A., Liu, C. C., Jensen, M. K. and Keasling, J. D. (2021), 'Integrating continuous hypermutation with high-throughput screening for optimization of cis, cis-muconic acid production in yeast', *Microbial biotechnology* **14**(6), 2617–2626. Publisher: Wiley Online Library.

Jiang, L., Dong, C., Liu, T., Shi, Y., Wang, H., Tao, Z., Liang, Y. and Lian, J. (2021), 'Improved functional expression of cytochrome p450s in saccharomyces cerevisiae through screening a cdna library from arabidopsis thaliana', *Frontiers in Bioengineering and Biotechnology* **9**, 764851.

Jiang, L., Huang, L., Cai, J., Xu, Z. and Lian, J. (2021), 'Functional expression of eukaryotic cytochrome p450s in yeast', *Biotechnology and Bioengineering* **118**(3), 1050–1065.

Jones, T. S., Oliveira, S. M., Myers, C. J., Voigt, C. A. and Densmore, D. (2022), 'Genetic circuit design automation with cello 2.0', *Nature protocols* **17**(4), 1097–1113.

Jorgensen, P., Nelson, B., Robinson, M. D., Chen, Y., Andrews, B., Tyers, M. and Boone, C. (2002), 'High-resolution genetic mapping with ordered arrays of saccharomyces cerevisiae deletion mutants', *Genetics* **162**(3), 1091–1099.

Kaczmarek, J. A. and Prather, K. L. (2021), 'Effective use of biosensors for high-throughput library screening for metabolite production', *Journal of Industrial Microbiology and Biotechnology* **48**(9-10), kuab049.

Kauffman, S. and Levin, S. (1987), 'Towards a general theory of adaptive walks on rugged landscapes', *Journal of theoretical Biology* **128**(1), 11–45.

Khalil, I., Quintens, G., Junkers, T. and Dusselier, M. (2020), 'Muconic acid isomers as platform chemicals and monomers in the biobased economy', *Green Chemistry* **22**(5), 1517–1541.

Khan, N., Yeung, E., Farris, Y., Fansler, S. J. and Bernstein, H. C. (2020), 'A broad-host-range event detector: expanding and quantifying performance between Escherichia coli and Pseudomonas species', *Synthetic Biology* **5**(1), ysaa002. Publisher: Oxford University Press.

Kitney, R. and Freemont, P. (2012), 'Synthetic biology–the state of play', *FEBS letters* **586**(15), 2029–2036.

Koch, M., Pandi, A., Borkowski, O., Batista, A. C. and Faulon, J.-L. (2019), 'Custom-made transcriptional biosensors for metabolic engineering', *Current Opinion in Biotechnology* **59**, 78–84. Publisher: Elsevier.

Kofoed, M., Milbury, K. L., Chiang, J. H., Sinha, S., Ben-Aroya, S., Giaever, G., Nislow, C., Hieter, P. and Stirling, P. C. (2015), 'An updated collection of sequence barcoded temperature-sensitive alleles of yeast essential genes', *G3: Genes, Genomes, Genetics* **5**(9), 1879–1887.

Kumar, A., Beloglazova, N., Bundalovic-Torma, C., Phanse, S., Deineko, V., Gagarinova, A., Musso, G., Vlasblom, J., Lemak, S., Hooshyar, M. et al. (2016), 'Conditional epistatic interaction maps reveal global functional rewiring of genome integrity pathways in escherichia coli', *Cell reports* **14**(3), 648–661.

Kuzmin, E., Sharifpoor, S., Baryshnikova, A., Costanzo, M., Myers, C. L., Andrews, B. J. and Boone, C. (2014), Synthetic genetic array analysis for global mapping of genetic networks in yeast, *in* 'Yeast Genetics', Springer, pp. 143–168.

Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Usaj, M. M., Van Leeuwen, J. et al. (2018), 'Systematic analysis of complex genetic interactions', *Science* **360**(6386), eaao1729.

Leavitt, J. M., Wagner, J. M., Tu, C. C., Tong, A., Liu, Y. and Alper, H. S. (2017), 'Biosensor-enabled directed evolution to improve muconic acid production in saccharomyces cerevisiae', *Biotechnology journal* **12**(10), 1600687.

Lennen, R. M., Lim, H. G., Jensen, K., Mohammed, E. T., Phaneuf, P. V., Noh, M. H., Malla, S., Börner, R. A., Chekina, K., Özdemir, E. et al. (2023), 'Laboratory evolution reveals general and specific tolerance mechanisms for commodity chemicals', *Metabolic Engineering* **76**, 179–192.

Levskaya, A., Weiner, O. D., Lim, W. A. and Voigt, C. A. (2009), 'Spatiotemporal control of cell signalling using a light-switchable protein interaction', *Nature* **461**(7266), 997–1001.

Lian, J., HamediRad, M., Hu, S. and Zhao, H. (2017), 'Combinatorial metabolic engineering using an orthogonal tri-functional crispr system', *Nature communications* **8**(1), 1–9.

Lian, J., Schultz, C., Cao, M., HamediRad, M. and Zhao, H. (2019), 'Multi-functional genome-wide crispr system for high throughput genotype–phenotype mapping', *Nature Communications* **10**.

Liu, C. C., Jewett, M. C., Chin, J. W. and Voigt, C. A. (2018), 'Toward an orthogonal central dogma', *Nature chemical biology* **14**(2), 103–106. Publisher: Nature Publishing Group US New York.

Lou, C., Stanton, B., Chen, Y.-J., Munsky, B. and Voigt, C. A. (2012), 'Ribozyme-based insulator parts buffer synthetic circuits from genetic context', *Nature biotechnology* **30**(11), 1137–1142.

Ma, J., Yu, M. K., Fong, S., Ono, K., Sage, E., Demchak, B., Sharan, R. and Ideker, T. (2018), 'Using deep learning to model the hierarchical structure and function of a cell', *Nature methods* **15**(4), 290.

Mancera, E., Bourgon, R., Brozzi, A., Huber, W. and Steinmetz, L. M. (2008), 'High-resolution mapping of meiotic crossovers and non-crossovers in yeast', *Nature* **454**(7203), 479–485.

Mani, R., Onge, R. P. S., Hartman, J. L., Giaever, G. and Roth, F. P. (2008), 'Defining genetic interaction', *Proceedings of the National Academy of Sciences* **105**(9), 3461–3466.

Mans, R., Daran, J.-M. G. and Pronk, J. T. (2018), 'Under pressure: evolutionary engineering of yeast strains for improved performance in fuels and chemicals production', *Current opinion in biotechnology* **50**, 47–56.

Martin, H., Shales, M., Fernandez-Piñar, P., Wei, P., Molina, M., Fiedler, D., Shokat, K. M., Beltrao, P., Lim, W. and Krogan, N. J. (2015), 'Differential genetic interactions of yeast stress response mapk pathways', *Molecular systems biology* **11**(4).

Mavrommati, M., Daskalaki, A., Papanikolaou, S. and Aggelis, G. (2022), 'Adaptive laboratory evolution principles and applications in industrial biotechnology', *Biotechnology Advances* **54**, 107795.

McCarty, N. S., Graham, A. E., Studená, L. and Ledesma-Amaro, R. (2020), 'Multiplexed crispr technologies for gene editing and transcriptional regulation', *Nature Communications* **11**(1), 1–13.

Meng, F. and Ellis, T. (2020), 'The second decade of synthetic biology: 2010–2020', *Nature Communications* **11**(1), 5174.

Meyer, A. J., Ellefson, J. W. and Ellington, A. D. (2015), 'Directed evolution of a panel of orthogonal t7 rna polymerase variants for in vivo or in vitro synthetic circuitry', *ACS synthetic biology* **4**(10), 1070–1076.

Mizoguchi, H., Sawano, Y., Kato, J.-i. and Mori, H. (2008), 'Superpositioning of deletions promotes growth of escherichia coli with a reduced genome', *DNA research* **15**(5), 277–284.

Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. and Voigt, C. A. (2012), 'Genetic programs constructed from layered logic gates in single cells', *Nature* **491**(7423), 249–253.

Moore, J. K. and Wethekam, L. (2021), 'Specialist α-tubulins for pluralist microtubules', *The Journal of Cell Biology* **220**(12).

Mullis, M. N., Matsui, T., Schell, R., Foree, R. and Ehrenreich, I. M. (2018), 'The complex underpinnings of genetic background effects', *Nature communications* **9**(1), 1–10.

Nielsen, A. A., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D. and Voigt, C. A. (2016), 'Genetic circuit design automation', *Science* **352**(6281), aac7341.

Nsamba, E. T., Bera, A., Costanzo, M., Boone, C. and Gupta Jr, M. L. (2021), 'Tubulin isotypes optimize distinct spindle positioning mechanisms during yeast mitosis', *Journal of Cell Biology* **220**(12), e202010155.

Olsson, H. (2020), 'Development of cri-spa, a mating-based, crispr-cas9 assisted method for high-throughput yeast strain construction, and its applications in yeast cell factory research', *PhD Thesis, DTU* .

Orelle, C., Carlson, E. D., Szal, T., Florin, T., Jewett, M. C. and Mankin, A. S. (2015), 'Protein synthesis by ribosomes with tethered subunits', *Nature* **524**(7563), 119–124.

Özaydın, B., Burd, H., Lee, T. S. and Keasling, J. D. (2013), 'Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production', *Metabolic engineering* **15**, 174–183.

Pan, X., Ye, P., Yuan, D. S., Wang, X., Bader, J. S. and Boeke, J. D. (2006), 'A dna integrity network in the yeast saccharomyces cerevisiae', *Cell* **124**(5), 1069–1081.

Pan, X., Yuan, D. S., Xiang, D., Wang, X., Sookhai-Mahadeo, S., Bader, J. S., Hieter, P., Spencer, F. and Boeke, J. D. (2004), 'A robust toolkit for functional profiling of the yeast genome', *Molecular cell* **16**(3), 487–496.

Payer, S. E., Marshall, S. A., Bärland, N., Sheng, X., Reiter, T., Dordic, A., Steinkellner, G., Wuensch, C., Kaltwasser, S., Fisher, K. and others (2017), 'Regioselective para-carboxylation of catechols with a prenylated flavin dependent decarboxylase', *Angewandte Chemie International Edition* **56**(44), 13893–13897. Publisher: Wiley Online Library.

Pena-Castillo, L. and Hughes, T. R. (2007), 'Why are there still over 1000 uncharacterized yeast genes?', *Genetics* **176**(1), 7–14.

Phaneuf, P. V., Gosting, D., Palsson, B. O. and Feist, A. M. (2019), 'Aledb 1.0: a database of mutations from adaptive laboratory evolution experimentation', *Nucleic acids research* **47**(D1), D1164–D1171.

Phillips, P. C. (2008), 'Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems', *Nature Reviews Genetics* **9**(11), 855–867.

Plekhova, V., Van Meulebroek, L., De Graeve, M., Perdones-Montero, A., De Spiegeleer, M., De Paepe, E., Van de Walle, E., Takats, Z., Cameron, S. J. and Vanhaecke, L. (2021), 'Rapid ex vivo molecular fingerprinting of biofluids using laser-assisted rapid evaporative ionization mass spectrometry', *Nature Protocols* **16**(9), 4327–4354.

Pokusaeva, V. O., Usmanova, D. R., Putintseva, E. V., Espinar, L., Sarkisyan, K. S., Mishin, A. S., Bogatyreva, N. S., Ivankov, D. N., Akopyan, A. V., Avvakumov, S. Y. et al. (2019), 'An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape', *PLoS genetics* **15**(4), e1008079.

Polturak, G. and Aharoni, A. (2018), '"la vie en rose": biosynthesis, sources, and applications of betalain pigments', *Molecular plant* **11**(1), 7–22.

Puchta, O., Cseke, B., Czaja, H., Tollervey, D., Sanguinetti, G. and Kudla, G. (2016), 'Network of epistatic interactions within a yeast snorna', *Science* **352**(6287), 840–844.

Pyne, M. E., Narcross, L., Melgar, M., Kevvai, K., Mookerjee, S., Leite, G. B. and Martin, V. J. (2018), 'An engineered aro1 protein degradation approach for increased cis, cis-muconic acid biosynthesis in saccharomyces cerevisiae', *Applied and environmental microbiology* **84**(17), e01095–18.

Qian, Y., Huang, H.-H., Jiménez, J. I. and Del Vecchio, D. (2017), 'Resource competition shapes the response of genetic circuits', *ACS synthetic biology* **6**(7), 1263–1272.

Rackham, O. and Chin, J. W. (2005), 'Cellular logic with orthogonal ribosomes', *Journal of the American Chemical Society* **127**(50), 17584–17585.

Rai, S. N., Singh, P., Varshney, R., Chaturvedi, V. K., Vamanu, E., Singh, M. and Singh, B. K. (2021), 'Promising drug targets and associated therapeutic interventions in parkinson's disease', *Neural regeneration research* **16**(9), 1730.

Raman, S., Rogers, J. K., Taylor, N. D. and Church, G. M. (2014), 'Evolution-guided optimization of biosynthetic pathways', *Proceedings of the National Academy of Sciences* **111**(50), 17803–17808.

Ravikumar, A., Arzumanyan, G. A., Obadi, M. K., Javanpour, A. A. and Liu, C. C. (2018), 'Scalable, continuous evolution of genes at mutation rates above genomic error thresholds', *Cell* **175**(7), 1946–1957.

Reid, R. J., González-Barrera, S., Sunjevaric, I., Alvaro, D., Ciccone, S., Wagner, M. and Rothstein, R. (2011), 'Selective ploidy ablation, a high-throughput plasmid transfer protocol, identifies new genes affecting topoisomerase i–induced dna damage', *Genome research* **21**(3), 477–486.

Reid, R. J., Sunjevaric, I., Voth, W. P., Ciccone, S., Du, W., Olsen, A. E., Stillman, D. J. and Rothstein, R. (2008), 'Chromosome-scale genetic mapping using a set of 16 conditionally stable saccharomyces cerevisiae chromosomes', *Genetics* **180**(4), 1799–1808.

Robertson, W. E., Funke, L. F., de la Torre, D., Fredens, J., Elliott, T. S., Spinck, M., Christova, Y., Cervettini, D., Böge, F. L., Liu, K. C. et al. (2021), 'Sense codon reassignment enables viral resistance and encoded polymer synthesis', *Science* **372**(6546), 1057–1062.

Roy, K. R., Smith, J. D., Vonesch, S. C., Lin, G., Tu, C. S., Lederer, A. R., Chu, A., Suresh, S., Nguyen, M., Horecka, J. et al. (2018), 'Multiplexed precision genome editing with trackable genomic barcodes in yeast', *Nature biotechnology* **36**(6), 512–520.

Rugbjerg, P., Sarup-Lytzen, K., Nagy, M. and Sommer, M. O. A. (2018), 'Synthetic addiction extends the productive life time of engineered escherichia coli populations', *Proceedings of the National Academy of Sciences* **115**(10), 2347–2352.

Ryan, O. W., Skerker, J. M., Maurer, M. J., Li, X., Tsai, J. C., Poddar, S., Lee, M. E., DeLoache, W., Dueber, J. E., Arkin, A. P. et al. (2014), 'Selection of chromosomal dna libraries using a multiplex crispr system', *Elife* **3**, e03703.

Sadhu, M. J., Bloom, J. S., Day, L., Siegel, J. J., Kosuri, S. and Kruglyak, L. (2018), 'Highly parallel genome variant engineering with crispr–cas9', *Nature genetics* **50**(4), 510–514.

Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O. and Feist, A. M. (2019), 'The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology', *Metabolic Engineering* **56**, 1–16.

Sasaki, N., Abe, Y., Goda, Y., Adachi, T., Kasahara, K. and Ozeki, Y. (2009), 'Detection of dopa 4, 5-dioxygenase (dod) activity using recombinant protein prepared from escherichia coli cells harboring cdna encoding dod from mirabilis jalapa', *Plant and Cell Physiology* **50**(5), 1012–1016.

Savitskaya, J., Protzko, R. J., Li, F.-Z., Arkin, A. P. and Dueber, J. E. (2019), 'Iterative screening methodology enables isolation of strains with improved properties for a FACS-based screen and increased L-DOPA production', *Scientific reports* **9**(1), 5815. Publisher: Nature Publishing Group UK London.

Schatz, P. J., Solomon, F. and Botstein, D. (1986), 'Genetically essential and nonessential alpha-tubulin genes specify functionally interchangeable proteins', *Molecular and cellular biology* .

Schuldiner, M., Collins, S. R., Thompson, N. J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J. F. et al. (2005), 'Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile', *Cell* **123**(3), 507–519.

Scott, M. and Hwa, T. (2023), 'Shaping bacterial gene expression by physiological and proteome allocation constraints', *Nature Reviews Microbiology* **21**(5), 327–342.

Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D. and Voigt, C. A. (2014), 'A 'resource allocator'for transcription based on a highly fragmented t7 rna polymerase', *Molecular systems biology* **10**(7), 742.

Segall-Shapiro, T. H., Sontag, E. D. and Voigt, C. A. (2018), 'Engineered promoters enable constant gene expression at any copy number in bacteria', *Nature biotechnology* **36**(4), 352–358.

Sharon, E., Chen, S.-A. A., Khosla, N. M., Smith, J. D., Pritchard, J. K. and Fraser, H. B. (2018), 'Functional genetic variants revealed by massively parallel precise genome editing', *Cell* **175**(2), 544–557.

Shepelin, D., Hansen, A. S. L., Lennen, R., Luo, H. and Herrgård, M. J. (2018), 'Selecting the best: evolutionary engineering of chemical production in microbes', *Genes* **9**(5), 249.

Shopera, T., He, L., Oyetunde, T., Tang, Y. J. and Moon, T. S. (2017), 'Decoupling resource-coupled gene expression in living cells', *ACS synthetic biology* **6**(8), 1596–1604. Publisher: ACS Publications.

Si, T., Luo, Y., Bao, Z. and Zhao, H. (2015), 'Rnai-assisted genome evolution in saccharomyces cerevisiae for complex phenotype engineering', *ACS synthetic biology* **4**(3), 283–291.

Siu, Y., Fenno, J., Lindle, J. M. and Dunlop, M. J. (2018), 'Design and selection of a synthetic feedback loop for optimizing biofuel tolerance', *ACS synthetic biology* **7**(1), 16–23.

Skjoedt, M. L., Snoek, T., Kildegaard, K. R., Arsovska, D., Eichenberger, M., Goedecke, T. J., Rajkumar, A. S., Zhang, J., Kristensen, M., Lehka, B. J. and others (2016), 'Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast', *Nature chemical biology* **12**(11), 951–958. Publisher: Nature Publishing Group US New York.

Skjoedt, M. L., Snoek, T., Kildegaard, K. R., Arsovska, D., Eichenberger, M., Goedecke, T. J., Rajkumar, A. S., Zhang, J., Kristensen, M., Lehka, B. J. et al. (2016), 'Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast', *Nature chemical biology* **12**(11), 951.

Snitkin, E. S., Dudley, A. M., Janse, D. M., Wong, K., Church, G. M. and Segrè, D. (2008), 'Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions', *Genome biology* **9**(9), R140.

Snoek, T., Chaberski, E. K., Ambri, F., Kol, S., Bjørn, S. P., Pang, B., Barajas, J. F., Welner, D. H., Jensen, M. K. and Keasling, J. D. (2020), 'Evolution-guided engineering of small-molecule biosensors', *Nucleic acids research* **48**(1), e3–e3. Publisher: Oxford University Press.

Snoek, T., Romero-Suarez, D., Zhang, J., Ambri, F., Skjoedt, M. L., Sudarsan, S., Jensen, M. K. and Keasling, J. D. (2018), 'An orthogonal and ph-tunable sensor-selector for muconic acid biosynthesis in yeast', *ACS synthetic biology* **7**(4), 995–1003.

Strucko, T., Magdenoska, O. and Mortensen, U. H. (2015), 'Benchmarking two commonly used Saccharomyces cerevisiae strains for heterologous vanillin--glucoside production', *Metabolic Engineering Communications* **2**, 99–108. Publisher: Elsevier.

Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M. J., Jelasity, M., Myers, C. L. et al. (2011), 'An integrated approach to characterize genetic interaction networks in yeast metabolism', *Nature genetics* **43**(7), 656–662.

Tabor, J. J., Salis, H. M., Simpson, Z. B., Chevalier, A. A., Levskaya, A., Marcotte, E. M., Voigt, C. A. and Ellington, A. D. (2009), 'A synthetic genetic edge detection program', *Cell* **137**(7), 1272–1281.

Tamsir, A., Tabor, J. J. and Voigt, C. A. (2011), 'Robust multicellular computing using genetically encoded nor gates and chemical 'wires'', *Nature* **469**(7329), 212–215.

Tang, H., Zhang, P. and Luo, X. (2022), 'Recent technologies for genetic code expansion and their implications on synthetic biology applications', *Journal of Molecular Biology* **434**(8), 167382.

Tas, H., Grozinger, L., Stoof, R., de Lorenzo, V. and Goñi-Moreno, (2021), 'Contextual dependencies expand the re-usability of genetic inverters', *Nature communications* **12**(1), 355. Publisher: Nature Publishing Group UK London.

Taylor, J. S. and Raes, J. (2004), 'Duplication and divergence: the evolution of new genes and old ideas', *Annu. Rev. Genet.* **38**, 615–643.

Tenaillon, O., Barrick, J. E., Ribeck, N., Deatherage, D. E., Blanchard, J. L., Dasgupta, A., Wu, G. C., Wielgoss, S., Cruveiller, S., Médigue, C. et al. (2016), 'Tempo and mode of genome evolution in a 50,000-generation experiment', *Nature* **536**(7615), 165–170.

Thompson, B., Pugh, S., Machas, M. and Nielsen, D. R. (2018), 'Muconic acid production via alternative pathways and a synthetic "metabolic funnel"', *ACS Synthetic Biology* **7**(2), 565–575.

Tong, A. H. Y., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C. W., Bussey, H. et al. (2001), 'Systematic genetic analysis with ordered arrays of yeast deletion mutants', *Science* **294**(5550), 2364–2368.

Voigt, C. A. (2020), 'Synthetic biology 2020–2030: six commercially-available products that are changing our world', *Nature Communications* **11**(1), 6379.

Wagih, O., Usaj, M., Baryshnikova, A., VanderSluis, B., Kuzmin, E., Costanzo, M., Myers, C. L., Andrews, B. J., Boone, C. M. and Parts, L. (2013), 'SGAtools: one-stop analysis and visualization of array-based genetic interaction screens', *Nucleic acids research* **41**(W1), W591–W596. Publisher: Oxford University Press.

Wang, G., Møller-Hansen, I., Babaei, M., D'Ambrosio, V., Christensen, H. B., Darbani, B., Jensen, M. K. and Borodina, I. (2021), 'Transportome-wide engineering of Saccharomyces cerevisiae', *Metabolic Engineering* **64**, 52–63. Publisher: Elsevier.

Wang, G., Øzmerih, S., Guerreiro, R., Meireles, A. C., Carolas, A., Milne, N., Jensen, M. K., Ferreira, B. S. and Borodina, I. (2020), 'Improvement of cis, cis-muconic acid production in Saccharomyces cerevisiae through biosensor-aided genome engineering', *ACS synthetic biology* **9**(3), 634–646. Publisher: ACS Publications.

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R. and Church, G. M. (2009), 'Programming cells by multiplex genome engineering and accelerated evolution', *Nature* **460**(7257), 894–898.

Wannier, T. M., Ciaccia, P. N., Ellington, A. D., Filsinger, G. T., Isaacs, F. J., Javanmardi, K., Jones, M. A., Kunjapur, A. M., Nyerges, A., Pal, C. et al. (2021), 'Recombineering and mage', *Nature Reviews Methods Primers* **1**(1), 7.

Weber, C., Brückner, C., Weinreb, S., Lehr, C., Essl, C. and Boles, E. (2012), 'Biosynthesis of cis, cis-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by Saccharomyces cerevisiae', *Applied and environmental microbiology* **78**(23), 8421–8430. Publisher: Am Soc Microbiol.

Weber, H. E., Gottardi, M., Brückner, C., Oreb, M., Boles, E. and Tripp, J. (2017), 'Requirement of a functional flavin mononucleotide prenyltransferase for the activity of a bacterial decarboxylase in a heterologous muconic acid pathway in Saccharomyces cerevisiae', *Applied and Environmental Microbiology* **83**(10), e03472–16. Publisher: Am Soc Microbiol.

Weill, U., Yofe, I., Sass, E., Stynen, B., Davidi, D., Natarajan, J., Ben-Menachem, R., Avihou, Z., Goldman, O., Harpaz, N. and others (2018), 'Genome-wide SWAp-Tag yeast libraries for proteome exploration', *Nature methods* **15**(8), 617–622. Publisher: Nature Publishing Group US New York.

Weiße, A. Y., Oyarzún, D. A., Danos, V. and Swain, P. S. (2015), 'Mechanistic links between cellular trade-offs, gene expression, and growth', *Proceedings of the National Academy of Sciences* **112**(9), E1038–E1047.

Williams, T. C., Pretorius, I. S. and Paulsen, I. T. (2016), 'Synthetic evolution of metabolic productivity using biosensors', *Trends in biotechnology* **34**(5), 371–381.

Wolfe, K. H. (2015), 'Origin of the yeast whole-genome duplication', *PLoS biology* **13**(8), e1002221.

Wright, S. et al. (1932), 'The roles of mutation, inbreeding, crossbreeding, and selection in evolution'.

Xiao, Y., Bowen, C. H., Liu, D. and Zhang, F. (2016), 'Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis', *Nature chemical biology* **12**(5), 339–344.

Xie, N.-Z., Liang, H., Huang, R.-B. and Xu, P. (2014), 'Biotechnological production of muconic acid: current status and future prospects', *Biotechnology advances* **32**(3), 615–622.

Xu, X., Meier, F., Blount, B. A., Pretorius, I. S., Ellis, T., Paulsen, I. T. and Williams, T. C. (2023), 'Trimming the genomic fat: minimising and re-functionalising genomes using synthetic biology', *Nature Communications* **14**(1), 1984.

Yachie, N., Petsalaki, E., Mellor, J. C., Weile, J., Jacob, Y., Verby, M., Ozturk, S. B., Li, S., Cote, A. G., Mosca, R. et al. (2016), 'Pooled-matrix protein interaction screens using barcode fusion genetics', *Molecular systems biology* **12**(4).

Yang, D., Kim, W. J., Yoo, S. M., Choi, J. H., Ha, S. H., Lee, M. H. and Lee, S. Y. (2018), 'Repurposing type iii polyketide synthase as a malonyl-coa biosensor for metabolic engineering in bacteria', *Proceedings of the National Academy of Sciences* **115**(40), 9835–9844.

Yilmaz, S., Nyerges, A., van der Oost, J., Church, G. M. and Claassens, N. J. (2022), 'Towards next-generation cell factories by rational genome-scale engineering', *Nature Catalysis* **5**(9), 751–765.

Yofe, I., Weill, U., Meurer, M., Chuartzman, S., Zalckvar, E., Goldman, O., Ben-Dor, S., Schütze, C., Wiedemann, N., Knop, M. and others (2016), 'One library to make them all: streamlining the creation of yeast libraries via a SWAp-Tag strategy', *Nature methods* **13**(4), 371–378. Publisher: Nature Publishing Group US New York.

Yu, M. K., Kramer, M., Dutkowski, J., Srivas, R., Licon, K., Kreisberg, J. F., Ng, C. T., Krogan, N., Sharan, R. and Ideker, T. (2016), 'Translation of genotype to phenotype by a hierarchy of cell subsystems', *Cell systems* **2**(2), 77–88.

Zhang, Y., Wang, J., Wang, Z., Zhang, Y., Shi, S., Nielsen, J. and Liu, Z. (2019), 'A grna-trna array for crispr-cas9 based rapid multiplexed genome editing in saccharomyces cerevisiae', *Nature communications* **10**(1), 1–10.

Chapter 2

Predicting genetic interactions with machine Learning

Foreword

Since the combinatorial space is large and genetic interactions are rare, their identification requires the mean to predict them computationally or to screen for them experimentally at high throughput. At the beginning of this PhD project, I attempted to use machine learning to learn the effects of gene double KO on fitness. As described in this chapter, I reimplemented DCell, the state of the art computational tool to predict genetic interaction of the SGA dataset produced by the Boone and Andrews Lab. If successful, such an approach could be presumably repeated on a different dataset to predict the effects of genetic interactions on a new phenotype such as productivity. However, this approach was unsuccessful as learning genetic interactions was complicated by data imbalance and noise which resulted in low predictive performances.

2.1 Abstract

The double mutant dataset released by the Boone and Andrews lab reports the occurrence of rare genetic interactions having a beneficial effect on fitness. This hints that positive genetic interactions could be exploited in a metabolic engineering context to improve strains traits such as fitness and productivity. However, due to dimensionality explosion, the systematic testing of the large number of possible combinations of genetic modifications is impractical in a routing metabolic engineering setting. Predicting positive genetic interactions would greatly reduce the number of combinations which need to be tested and would therefore be very advantageous. Here, I explore this possibility by attempting to reproduce DCell, the state of the art deep neural network trained to predict genetic interactions.

2.2 Introduction

In yeast genetics, a "genetic interaction" describes the occurrence of an unexpected phenotype arising from the combination of two genetic modifications. Because of its genetic tractability, yeast has been instrumental in the experimental study of genetic interactions. Famously, the Synthetic Gene Array (SGA) screen has enabled the quantification of the effect on fitness of close to all gene double knock-out (KO) in yeast. Briefly, SGA mates two haploid parent strains knocked-out at a different locus each marked with a selectable marker. The mated diploid is sporulated and grown on double-selective medium to co-segregate the two mutated loci in thehaploid progeny (Tong et al., 2001). From a proof-of-concept screening 8 non-essential genes, the availability of the full yeast knock out (YKO) library enabled the phenotyping of close to all possible yeast double mutants of both non-essential and essential genes (**?**). SGA also provides the data analysis framework to formalise the quantification of interactions (Baryshnikova et al., 2010). It defines a genetic interaction as the difference between the observed fitness of the double mutant and its expected fitness based on the fitnesses of its two individual mutants. SGA uses the multiplicative assumption (Mani et al., 2008), defining the expected fitness as the multiplication of the fitness of the

individual mutants **(Fig. 2.1a)**.

The systematic quantification of interactions for all double mutants made possible by SGA is a precious resource from which a number of observations can be drawn. Firstly, positive interactions are rare, representing only 1.2% of combinations in the Non-Essential x Non-Essential dataset (henceforth referred to as the double mutant dataset) and only 49% of these display an improved fitness compared to WT. However, positive interactions are responsible for the fittest possible strains in the landscape. The 30% increase in fitness observed in the fittest 3311 double mutants was expected based on single mutant fitness for only 3 strains. In other words, the 3308 fittest double mutants were generated by positive interactions unsuspected from the fitness of their single mutants **(Fig. 2.1b)**. Assuming that this observation holds true for other traits, this is of extraordinary relevance to metabolic engineers. Indeed, this hints that the interaction space holds synergistic phenotype maximas waiting to be exploited.

Yet another lesson from SGA is that systematically generating double mutants is extremely laborious and repeating this approach for a new phenotype in a metabolic engineering context (e.g. compound synthesis in a production strain) is highly impractical. Therefore, predicting positive interactions that would guide a metabolic engineer to the most productive strains would reduce the genetic space that needs to be tested.

A number of computational tools have been developed to predict the effects of genetic interactions on fitness. A successful approach in oncology is to mine cancer cell line datasets for patterns hinting to the presence of double mutants leading to cell death called synthetic lethals. For example, pairs of synthetic lethals can be recommended based on the observation that they are underrepresented in such datasets (Jerby-Arnon et al., 2014; Sinha et al., 2017). The use of metabolic models to recommend combinations of KO improving a product yield has also been reported (Burgard et al., 2003; Harrison et al., 2007; Barker et al., 2015). However their predictions have been shown to be highly inaccurate with a particularly low recall (Alzoubi et al., 2019; Szappanos et al., 2011). Moreover, the scope of metabolic model predictions is limited to metabolic genes failing to depict other important cellular functions. Finally, numerous supervised machine learning approaches have been trained to classify genetic interactions. More specifically, the task is to identify pairs of interacting genes based on input features specific to the gene pair and given a dataset of interacting and non-interacting gene pairs. A wide variety of data can be employed to create features characterising gene pairs (e.g. fusion, coexpression and physical interactions) that are used for training (Pandey et al., 2010; Wong et al., 2004; Yu et al., 2016; Benstead-Hume et al., 2019). A popular approach is to compile features characterising gene pairs from interaction networks such as the Gene Ontology (GO) or the STRING protein interactions (Consortium, 2006; Szklarczyk et al., 2021). For example, the "shortest path" or the "number of mutual neighbours" are two features which can be extracted for gene pairs in a network and used for training (Benstead-Hume et al., 2019). In another example, Yu et al. invented a feature called the "ontotype" which captures the impact one (or multiple) gene KO has on each term of the cell's hierarchy. For each term, the impact is defined as the number of daughter terms which are themselves impacted. A random forest classifier trained on the ontotypes of KO pairs was able to predict interactions between pairs of mutants with a Pearson's Correlation Coefficient (PCC) of 0.35 (Yu et al., 2016). This is a notable achievement given that repeats of SGA screens showed a PCC of 0.69 (Kuzmin et al., 2018).

In 2018, a deep neural network called DCell was reported to predict gene interactions on the double mutant dataset with a PCC of 0.5 (Ma et al., 2018). DCell has a particular architecture which reproduces the topology of the GO. The nodes in the network represent GO terms and the edges connecting nodes represent GO child-parent relationship. An advantage of DCell is that a genotype is encoded as a simple hot encoded vector. Each position in the vector represents a gene and a double mutant is encoded with a one at the two knocked-out gene positions while the rest of the gene positions are zeros **(Fig. 2.1c)**. This input is fed into DCell via neurons representing GO-gene annotations. The neural network propagates activation up its GO-like architecture and outputs a floating value between -1 and 1, predicting the interaction score of the two gene KO.

DCell shows state of the art predictions for genetic interactions. However, this required training on millions of data points generated by SGA (Ma et al., 2018). In a metabolic engineering context, obtaining a dataset covering millions of double mutant for a metric of productivity is unfeasible. In addition, the requirements to generate data on most of the space beats the purpose of obtaining a predictive tool since predictions are only useful for yet untested combinations. A standing question is therefore whether DCell can maintain some predictive power when trained on a fraction of the double mutant space. Here, I test the possibility of training DCell on a fraction of the double mutant dataset. I show that data imbalance critically impairs DCell performance. Finally, I rule out

the possibility of obtaining useful predictions from DCell when trained on an amount of data obtainable within a typical metabolic engineering project.

2.3 Materials and Methods

2.3.1 SGA and GO data.

This work used the Non essential x Non essential gene dataset published by Constanzo et al. (2016) downloaded from https://thecellmap.org/costanzo2016 . GO gene associations were extracted from gene_association.sgd.20170914.gz downloaded from
http://sgd-archive.yeastgenome.org/curation/literature/archive/. The GO ontologies were downloaded from http://geneontology ontology/

DCell architecture

The GO ontology is input as a list of connections between a parent term and its children or/and its associated genes. The gene ID map is the list of gene inputs to DCell (fig1). The scripts necessary to generate these files from the GO database were not provided by the authors. We therefore generated the GO ontology with the following procedure:

1. The gene ID map was generated by extracting all unique query and array genes present in the SGA None essential x None essential (NxN) gene dataset resulting in a list of 3597 genes.
2. A starting ontology storing GO terms relationships and GO term association with yeast genes were extracted from the GO and SGD databases, respectively.
3. Genes not present in the gene ID map were removed from the ontology ensuring that the input of DCell matches its training dataset.
4. We delete GO-term:gene associations from parents GO-terms if the same gene was associated with a children GO-term. This keeps GO-term:gene association to genes to the deepest term only as described (Ma et al., 2018)
5. GO-terms with less than 6 genes associated with themselves or their descent were deleted from the ontology. Note: this removed many terms irrelevant to yeast.
6. Redundant terms (i.e. terms which connect one parent term to one child term) were removed

Following these 6 steps, the processed ontology was saved as a text file matching the input format of DCell. In addition, the user can choose to create multiple nodes for each GO term within DCell so that each GO term connection is a n_node by n_node fully connected layer.

2.3.2 Availability

Scripts used in this chapter are available at: https://github.com/pc2912/DCell_Repoblic_online

2.4 Results

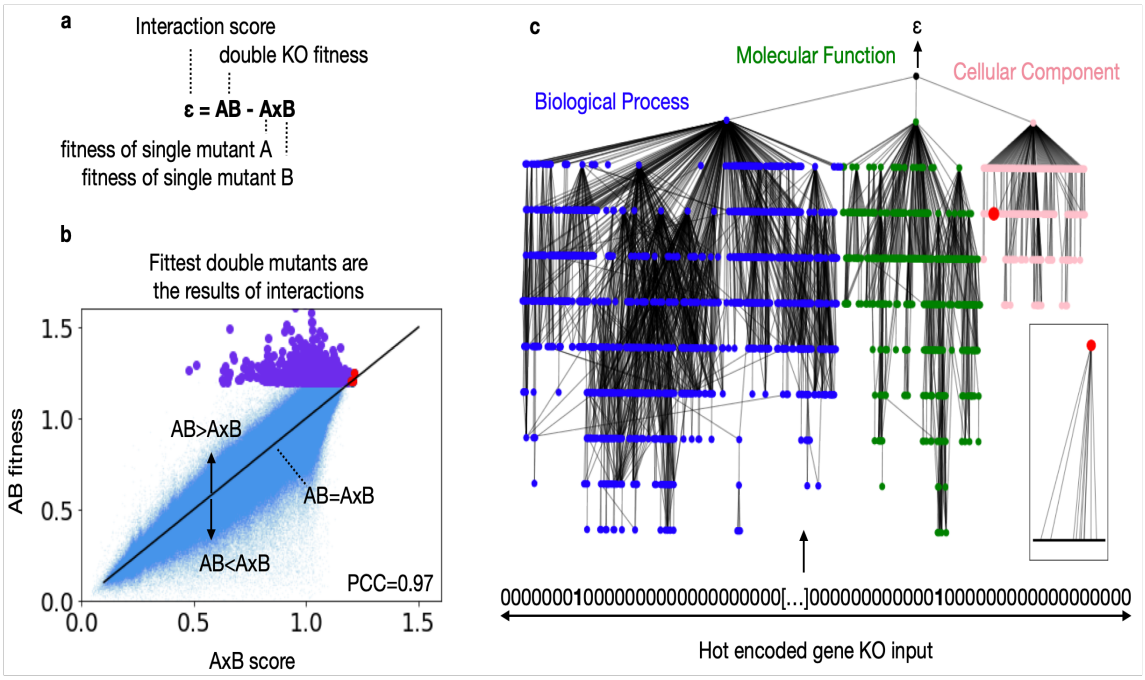


Figure 2.1: **The Double Mutant Dataset and DCell**
a) SGA defines the interaction score ϵ as the difference between the observed double mutant fitness AB and the fitness predicted by the multiplication of the fitness of single mutant A and of single mutant B. **b)** Correlation between the expected and the actual double mutant fitness. The double mutants falling above and under the AB=AxB line show positive and negative interactions, respectively. The 3311 double mutants with a 30% increase in fitness (purple) were only predictable in three cases (red) **c)** DCell architecture. Each node represents a GO term and edges are GO relationships. The input of DCell is a hot-encoded vector where each position represents a gene. Gene knock-outs are input as ones. Nodes are coloured according to their GO aspect (pink: cellular component, green: molecular function, blue: biological process). Gene to term connections (≈34 thousand) were omitted for visibility but are shown in the inset for the GO term "electron transport chain" GO:0022900 highlighted in red.

DCell was built using its python implementation released by Kuenzi et al. (Kuenzi et al., 2020). Briefly, the script requires two inputs: the GO and a gene ID map. The GO is used as a template when building DCell so that the neural connection of the network repeats the topology of the ontology. The gene ID map is the list of genes serving as inputs in DCell. The Ideker lab did not release the scripts necessary to process GO ontology files into the format required by DCell nor the exact datasets DCell was trained on. Training on the non-essential double mutant dataset, the gene ID map was defined as the set of query and array strains in this dataset. The GO was processed to match DCell input following the author's textual description (material and methods). This produced working version of DCell made of 1457 nodes (i.e. GO terms), 3587 input features (i.e. genes) and 132.7million trainable parameters **(Fig. 2.1c)**.

DCell was trained as described on 90% of the double mutant dataset with the Adam optimizer to reduce the mean square error loss (Kingma and Ba, 2014). The predictions obtained were notably poorer (PCC=0.12) than reported in the publication (PCC=0.5, **Fig. 2.2a and b**). Whatsmore, the distribution of the predictions was narrowly centred around 0 **(Fig. 2.2c)**. This is because the data is severely imbalanced with most double knock-outs having an interaction score close to 0 **(Fig. 2.2a)**. Therefore the model minimises the loss by predicting scores close to 0 for most double mutant.

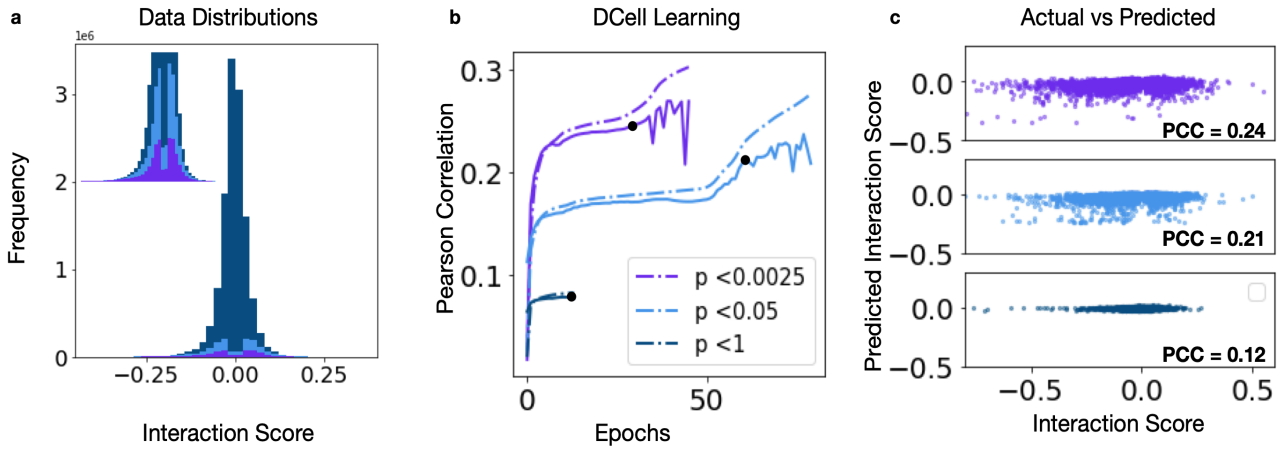


Figure 2.2: **Data imbalance challenges the learning of genetic interactions**
DCell was trained on the double mutant dataset filtered for increasingly stringent p-values (dark blue: p<1 (i.e. no filtering), light blue p<0.05, purple p<0.0025. **a)** Interaction score distributions after filtering. The unfiltered distribution has a strong imbalance with most double mutants showing no interaction. Filtering high p-values removes mutant with a low interaction score. **b)** DCell performances reported as the PCC on the training set (dashed lines) and test set (full lines) across epochs. Filtering the dataset helps DCell reach a higher PCC. The model trained up to the epoch displayed by the black dot were used to make the predictions in c. **c)** Prediction of 15,000 test-set double mutants by DCell trained on the differently filtered datasets showed in b.

In their publication, the DCell authors state that they trained the model on the 3 million or 8 million data points made available by Costanzo et al. 2010 and 2016. However, the double mutant datasets released by Costanzo et al. in 2010 and 2016 contain 6.6 and 12.7 million data points, respectively (**??**). It seems that the authors also filtered their dataset, although the nature of the filter was not disclosed. The data was filtered according to the p-value attached with each interaction score in the double mutant dataset. This reduced the data imbalance by removing most double knockouts with a low interaction score **(Fig. 2.2a)** which improved DCell learning **(Fig. 2.2b)**. Whatsmore, the higher the stringency of the filtering, the better was DCell able to learn. For example, removing data points with p>0.0025, DCell was able to reach a PCC of 0.25 **(Fig. 2.2b)**.

So far, the random split of the data into the training and the test set is not faithful to the practical constraints of SGA screens. Indeed, SGA crosses a query strain to the ≈4000 array strains of the yeast knockout library. Each query strain enables on screen which generates a batch of ≈ 4000 data points. Hence, if the full dataset is not yet available, all possible pair of double mutants are not equally accessible. Rather, the available training set is constrained by the identity of the query strains screened. A realistic way of splitting the data into a train and test set is therefore to apply the random split on the list of query strains and sort the accompanying batch of double mutants accordingly **(Fig. 2.3a)**. This 'query mode' split samples less query gene diversity should hence make training more difficult **(Fig. 2.3a)**. In addition, I previously trained DCell on 90% of the double mutant data. In practice, a model is valuable if it can learn from a fraction of the space to make predictions on the rest of the unknown double mutants. I prepared more realistic training conditions by randomly splitting the data in 'query mode' by selecting the double mutants associated with a set of 5, 50 or 1800 query strains. In this setting, 5, 50 or 1800 query strains and their associated double mutants were placed in the training set while the rest of the query strains and their associated double mutants were placed in the test set set. With the resources at hands for my PhD project, screening 5, 50 or 1800 query strains ranges from very realistic to possible to impossible. To help training, I balanced the data by including points with a p-value<0.05.

When trained on data generated by 5 and 50 screens DCell showed almost no learning with predictions on a test-set having a PCC close to 0. When trained on the data generated with 1800 screens, DCell showed a PCC of 0.13 **(Fig. 2.3b)**. Such poor predicting power would be of little help in an experimental setting. Indeed, when predicting 15,000 randomly drawn data points from the test-set, of the 100 predicted top double mutants, only two (p=0.11) to three (p=0.025) were also found in the actual top 100 **(Fig. 2.3c)**. This is only slightly better than randomly drawing double mutants.

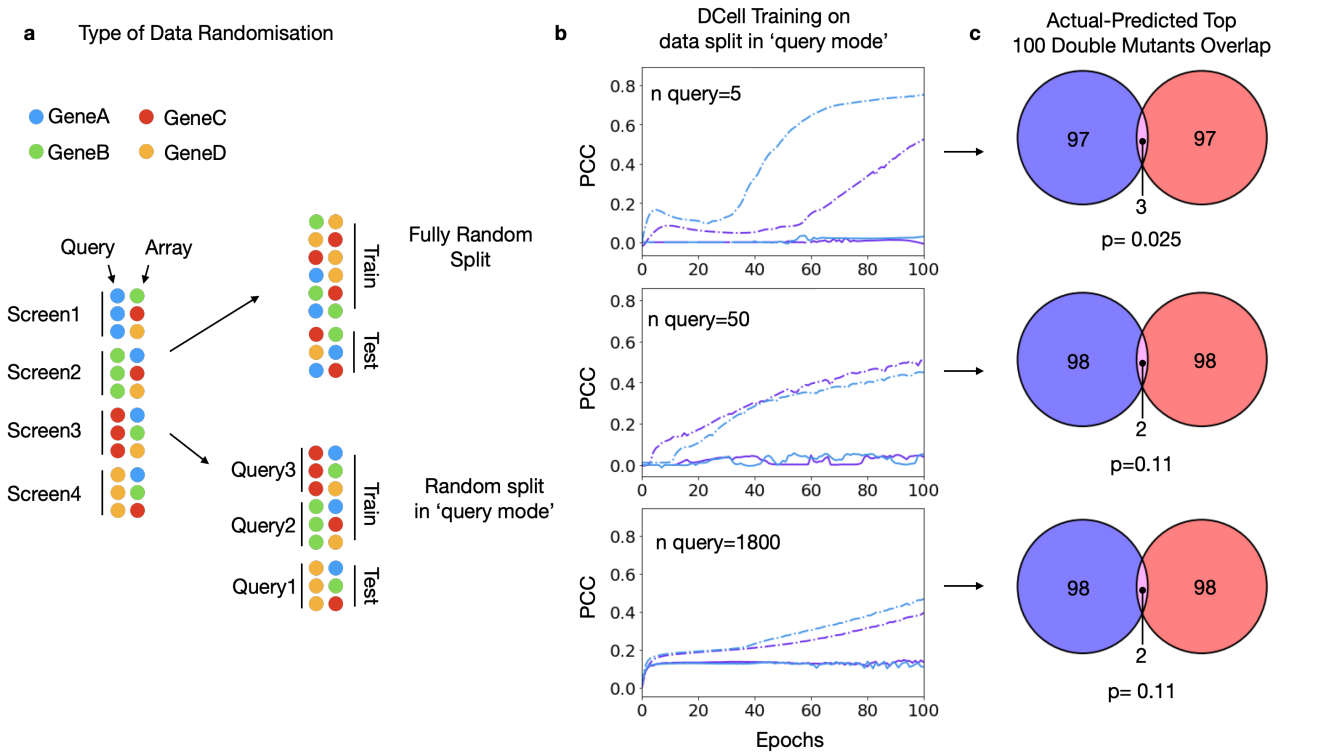


Figure 2.3: **Training DCell on a fraction of the double mutant space split in 'query mode'.** **a)** A fully random split has access to all double mutants in the dataset whereas splitting in 'query mode' randomly assigns query strains and their accompanying double mutants to the train or the test sets. **b)** DCell performances reported as the PCC on the training set (hashed lines) and test set (full lines) across epochs. The training sets were obtained by splitting the data in query mode and placing 5, 50 or 1800 query strains in the training set. Blue and purple lines show training on training sets made from different sets of query strains. **c)** Overlap between the top 100 strains predicted by the models trained in b) and the actual top strains for a set of 15,000 double mutants randomly drawn from the test set. the p-value is the probability of drawing k the number of observed overlapping mutants at random. This is given by the hypergeomtric distribution: $p(x \geq k|M, n, N)$. With M=15000, n=N=100.

2.5 Discussion

In this short chapter, DCell the state of the art neural network to predict genetic interactions was reproduced. In favourable training conditions, DCell reached half of the published predictive power, with a PCC of 0.25 on a cross validated dataset. The chief difficulty in learning genetic interactions is their rarity. With most gene pairs displaying close to no interaction, DCell minimised the loss function by predicting no interactions for all pairs. This data imbalance could be palliated by filtering the dataset on the p-value accompanying every double mutant data point. This removed most non-interacting gene pairs and helped DCell reach higher predictive ability. I note however that filtering the dataset based on a p-value threshold is not realistic in an experimental setting. Indeed, the p-value is computed based on the gene KO quadruplicate fitness score (**?**) . In other words, filtering on the p-value requires having access to the outcome of a double mutant before predicting it.

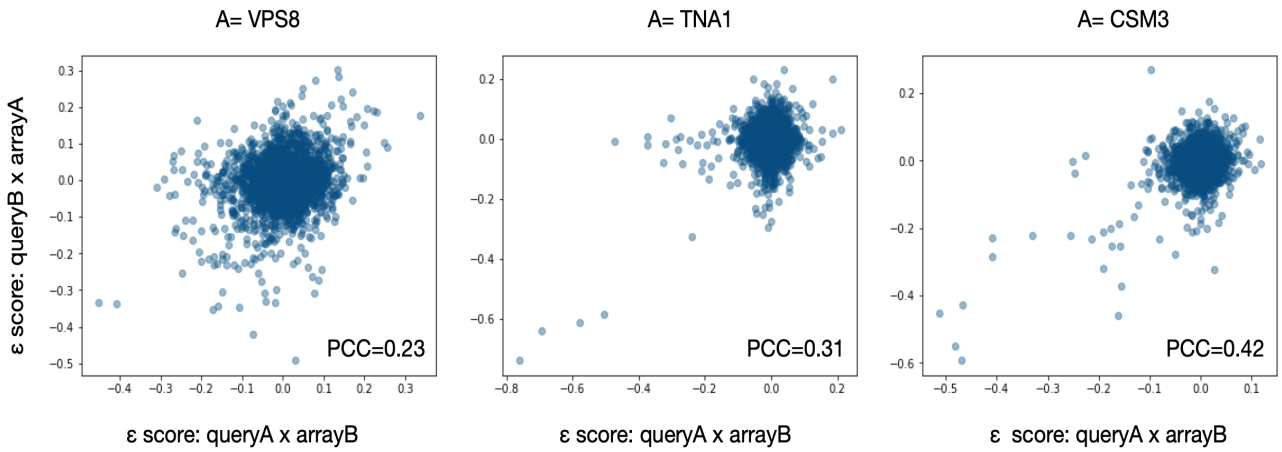


Figure 2.4: **Correlation of the interaction scores for the same double mutants obtained from reverse crosses in the double mutant dataset.**
The interaction score of double mutant AB can be measured by crossing query strain A with array strain B or query strain B with array strain A. The correlation for these replicates are shown for three query genes: left: *VPS8*, center: *TNA1*, right: *CSM3*.

Besides data imbalance, learning genetic interactions is probably also complicated by the noise in the dataset. SGA screens display a limited reproducibility for double mutant fitnesses with a PCC of 0.67-0.69 (Kuzmin et al., 2018; Jaffe et al., 2019). This reproducibility is calculated for double mutant fitness and becomes even lower for the double mutant interaction score. Indeed, interactions measured for the same double mutants but obtained from reverse crosses (i.e. query(A)xarray(B) vs query(B)xarray(A)) display a PCC of 0.3-0.4 **(Fig. 2.4)**. Given this upper limit, it is surprising that the authors of DCell reached a predictive PCC of 0.5. This high performance could not be reproduced given that the filtration step applied to the data prior to training was left undisclosed.

In addition to the interaction score, the authors also trained DCell to learn the fitness of the double mutants for which they also reported a PCC of 0.5. This is a second surprise since the multiplicative assumption already predicts double mutant fitness with a much higher PCC of 0.97 (figure 1b). When trained to learn fitness rather than interaction, DCell predictions produced a PCC close to 0.9 (not shown). Altogether, the results reported by Ma et al. are puzzling in light of the little information they provide on their data processing.

In contrast to experimental studies, computational methods have the luxury to be fully transferable and their publication should aim to achieve reproducibility within a few clicks and keystrokes (Heil et al., 2021). The design of DCell architecture and the compiling of its training data was only described textually in the publication. Contacting the Ideker lab to obtain further guidance was also unsuccessful. The lack of transparency over computational methods even in high impact journals is courageously being raised (Haibe-Kains et al., 2020) and according to a recently proposed reproducibility standard, Ma et al. would not reach the lowest "bronze" score (Heil et al., 2021). Further, despite the impact of the publication which now counts over 300 citations, I am not aware of its reproduction by other studies learning genetic interactions of synthetic lethality (Benstead-Hume et al., 2019; Cai et al., 2020; Gervits and Sharan, 2022). This is especially surprising in machine learning where state of the art models are routinely reproduced for benchmarking.

Beyond DCell, our results advocate for better transparency on the processing of the training data in machine learning. The performance of a learner goes hand-in-hand with the nature of the data used for training and validation. Yet, data preprocessing might alter its characteristics, helping a model perform better than on realistic data. The concept of volatility was introduced to quantify the effects of data pre-processing steps on the performance of a model (Zelaya, 2019). Reporting the volatility of a model is particularly relevant when training on biological data where data imbalance is especially severe (Chicco, 2017). For example, Benstead-Hume et al. artificially balanced their training set to train a tree classifying synthetic lethals. However, the authors also tested prediction performances when making predictions on an unbalanced set, supporting the robustness of their framework (Benstead-Hume et al., 2019).

To conclude, this attempt to predict interactions with the state of the art computational tool was unsuccessful. As a complementary approach, the remaining of this thesis, focuses on developing a fully experimental framework for screening the interaction space and uncovering beneficial interactions.

Bibliography

Alzoubi, D., Desouki, A. A. and Lercher, M. J. (2019), 'Flux balance analysis with or without molecular crowding fails to predict two thirds of experimentally observed epistasis in yeast', *Scientific reports* **9**(1), 1–9.

Barker, B., Xu, L. and Gu, Z. (2015), 'Dynamic epistasis under varying environmental perturbations', *PloS one* **10**(1).

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S. et al. (2010), 'Quantitative analysis of fitness and genetic interactions in yeast on a genome scale', *Nature methods* **7**(12), 1017.

Benstead-Hume, G., Chen, X., Hopkins, S. R., Lane, K. A., Downs, J. A. and Pearl, F. M. (2019), 'Predicting synthetic lethal interactions using conserved patterns in protein interaction networks', *PLoS computational biology* **15**(4), e1006888.

Burgard, A. P., Pharkya, P. and Maranas, C. D. (2003), 'Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization', *Biotechnology and bioengineering* **84**(6), 647–657.

Cai, R., Chen, X., Fang, Y., Wu, M. and Hao, Y. (2020), 'Dual-dropout graph convolutional network for predicting synthetic lethality in human cancers', *Bioinformatics* **36**(16), 4458–4465.

Chicco, D. (2017), 'Ten quick tips for machine learning in computational biology', *BioData mining* **10**(1), 35.

Consortium, G. O. (2006), 'The gene ontology (GO) project in 2006', *Nucleic acids research* **34**(suppl_1), D322–D326. Publisher: Oxford University Press.

Gervits, A. and Sharan, R. (2022), 'Predicting genetic interactions, cell line dependencies and drug sensitivities with variational graph auto-encoder', *Frontiers in Bioinformatics* **2**, 1025783.

Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., of Directors Shraddha Thakkar 35 Kusko Rebecca 36 Sansone Susanna-Assunta 37 Tong Weida 35 Wolfinger Russ D. 38 Mason Christopher E. 39 Jones Wendell 40 Dopazo Joaquin 41 Furlanello Cesare 42, M. A. Q. C. M. S. B., Waldron, L., Wang, B., McIntosh, C., Goldenberg, A., Kundaje, A. et al. (2020), 'Transparency and reproducibility in artificial intelligence', *Nature* **586**(7829), E14–E16.

Harrison, R., Papp, B., Pál, C., Oliver, S. G. and Delneri, D. (2007), 'Plasticity of genetic interactions in metabolic networks of yeast', *Proceedings of the National Academy of Sciences* **104**(7), 2307–2312.

Heil, B. J., Hoffman, M. M., Markowetz, F., Lee, S.-I., Greene, C. S. and Hicks, S. C. (2021), 'Reproducibility standards for machine learning in the life sciences', *Nature Methods* **18**(10), 1132–1135.

Jaffe, M., Dziulko, A., Smith, J. D., Onge, R. P. S., Levy, S. F. and Sherlock, G. (2019), 'Improved discovery of genetic interactions using crispriseq across multiple environments', *Genome research* **29**(4), 668–681.

Jerby-Arnon, L., Pfetzer, N., Waldman, Y. Y., McGarry, L., James, D., Shanks, E., Seashore-Ludlow, B., Weinstock, A., Geiger, T., Clemons, P. A. et al. (2014), 'Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality', *Cell* **158**(5), 1199–1209.

Kingma, D. P. and Ba, J. (2014), 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980* .

Kuenzi, B. M., Park, J., Fong, S. H., Sanchez, K. S., Lee, J., Kreisberg, J. F., Ma, J. and Ideker, T. (2020), 'Predicting drug response and synergy using a deep learning model of human cancer cells', *Cancer cell* **38**(5), 672–684.

Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Usaj, M. M., Van Leeuwen, J. et al. (2018), 'Systematic analysis of complex genetic interactions', *Science* **360**(6386), eaao1729.

Ma, J., Yu, M. K., Fong, S., Ono, K., Sage, E., Demchak, B., Sharan, R. and Ideker, T. (2018), 'Using deep learning to model the hierarchical structure and function of a cell', *Nature methods* **15**(4), 290.

Mani, R., Onge, R. P. S., Hartman, J. L., Giaever, G. and Roth, F. P. (2008), 'Defining genetic interaction', *Proceedings of the National Academy of Sciences* **105**(9), 3461–3466.

Pandey, G., Zhang, B., Chang, A. N., Myers, C. L., Zhu, J., Kumar, V. and Schadt, E. E. (2010), 'An integrative multi-network and multi-classifier approach to predict genetic interactions', *PLoS computational biology* **6**(9).

Sinha, S., Thomas, D., Chan, S., Gao, Y., Brunen, D., Torabi, D., Reinisch, A., Hernandez, D., Chan, A., Rankin, E. B. et al. (2017), 'Systematic discovery of mutation-specific synthetic lethals by mining pan-cancer human primary tumor data', *Nature communications* **8**(1), 15580.

Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M. J., Jelasity, M., Myers, C. L. et al. (2011), 'An integrated approach to characterize genetic interaction networks in yeast metabolism', *Nature genetics* **43**(7), 656–662.

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P. and others (2021), 'The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets', *Nucleic acids research* **49**(D1), D605–D612. Publisher: Oxford University Press.

Tong, A. H. Y., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C. W., Bussey, H. et al. (2001), 'Systematic genetic analysis with ordered arrays of yeast deletion mutants', *Science* **294**(5550), 2364–2368.

Wong, S. L., Zhang, L. V., Tong, A. H., Li, Z., Goldberg, D. S., King, O. D., Lesage, G., Vidal, M., Andrews, B., Bussey, H. et al. (2004), 'Combining biological networks to predict genetic interactions', *Proceedings of the National Academy of Sciences* **101**(44), 15682–15687.

Yu, M. K., Kramer, M., Dutkowski, J., Srivas, R., Licon, K., Kreisberg, J. F., Ng, C. T., Krogan, N., Sharan, R. and Ideker, T. (2016), 'Translation of genotype to phenotype by a hierarchy of cell subsystems', *Cell systems* **2**(2), 77–88.

Zelaya, C. V. G. (2019), Towards explaining the effects of data preprocessing on machine learning, *in* '2019 IEEE 35th international conference on data engineering (ICDE)', IEEE, pp. 2086–2090.

Chapter 3

Manuscript 1: CRI-SPA: a high-throughput method for systematic genetic editing of yeast libraries

Foreword

Having ruled out the possibility of guiding the search for genetic interactions with machine learning, I focused the rest of my studies on identifying them with the help of high throughput screen. As reviewed in Chapter 1, several screenings methods were available with different exploitation and exploration trade-offs. With its ability to systematically test all variants in a genetic library, SGA seemed particularly advantageous. Indeed, SGA tests the interaction of a genetic feature along every gene dimension of the host, generating the most complete interaction dataset. Luckily, a similar array screen named CRI-SPA was in development at DTU bioengineering and its development needed to be pursued.

When I overtook CRI-SPA, the experimental part of the screen was solidly in place thanks to the previous efforts of Helén Olson and Hilde Coumou during their doctoral studies. As a proof of concept, CRI-SPA had been developed to deliver the pathway responsible for the synthesis of the yellow plant pigment betaxanthin to the YKO library. The effect of the gene mutant in the library background on the betaxanthin pathway could be observed by the difference in colony yellowness for all mutants in the library. This still needed an image analysis workflow extracting yellow color from the arrays of edited colonies to formalise their comparison.

Overtaking the lead on CRI-SPA after the Helén's graduation, I contributed by developing this analysis workflow which revealed numerous host:pathway interactions and validated their impact on betaxanthin synthesis by reverse engineered some of the hits. This data took part in the first article describing CRI-SPA published in Nucleic Acids Research https://doi.org/10.1093/nar/gkad656 which is reproduced below with the authorisation of the journal.

Nucleic Acids Research, 2023 **1**
https://doi.org/10.1093/nar/gkad656

CRI-SPA: a high-throughput method for systematic genetic editing of yeast libraries

Paul Cachera [1,†], Helén Olsson [1,†], Hilde Coumou [2,†], Mads L. Jensen [2], Benjamín J. Sánchez [2], Tomas Strucko [2], Marcel van den Broek [3], Jean-Marc Daran [3], Michael K. Jensen [1], Nikolaus Sonnenschein [2], Michael Lisby [4] and Uffe H. Mortensen [2,*]

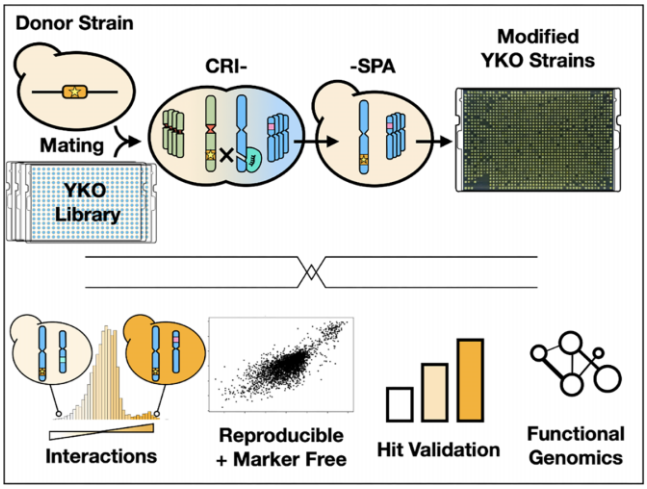
[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Denmark, [2]Department of Biotechnology and Biomedicine, Technical University of Denmark, Denmark, [3]Department of Biotechnology, Delft University of Technology, Delft, The Netherlands and [4]Department of Biology, University of Copenhagen, Copenhagen, Denmark

Received October 18, 2022; Revised July 07, 2023; Editorial Decision July 11, 2023; Accepted August 10, 2023

ABSTRACT

Biological functions are orchestrated by intricate networks of interacting genetic elements. Predicting the interaction landscape remains a challenge for systems biology and new research tools allowing simple and rapid mapping of sequence to function are desirable. Here, we describe CRI-SPA, a method allowing the transfer of chromosomal genetic features from a CRI-SPA Donor strain to arrayed strains in large libraries of *Saccharomyces cerevisiae*. CRI-SPA is based on mating, CRISPR-Cas9-induced gene conversion, and Selective Ploidy Ablation. CRI-SPA can be massively parallelized with automation and can be executed within a week. We demonstrate the power of CRI-SPA by transferring four genes that enable betaxanthin production into each strain of the yeast knockout collection (≈4800 strains). Using this setup, we show that CRI-SPA is highly efficient and reproducible, and even allows marker-free transfer of genetic features. Moreover, we validate a set of CRI-SPA hits by showing that their phenotypes correlate strongly with the phenotypes of the corresponding mutant strains recreated by reverse genetic engineering. Hence, our results provide a genome-wide overview of the genetic requirements for betaxanthin production. We envision that the simplicity, speed, and reliability offered by CRI-SPA will make it a versatile tool to forward systems-level understanding of biological processes.

GRAPHICAL ABSTRACT



INTRODUCTION

Baker's yeast *Saccharomyces cerevisiae* is an important biological model organism and a key production host in industrial biotechnology. *S. cerevisiae* was the first eukaryote to be fully sequenced (1,2), and its simple lifecycle as a single cell organism with a highly developed genetic toolbox has positioned this yeast as a frontrunner in the field of systems biology. However, the unpredictability of phenotypic effects resulting from combinations of different genetic traits still challenges our fundamental understanding of yeast and its engineering. The introduction of CRISPR for yeast engineering (3) promises to relieve this bottleneck by accelerating the systematic construction and testing of genetic variants. For example, in numerous genome-wide engineering projects, CRISPR-based strategies have been developed to create and screen libraries of strains (4–11).

*To whom correspondence should be addressed. Tel: +45 45 25 27 01; Fax: +45 4588 4922; Email: um@bio.dtu.dk
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

These screens generally follow a one-pot workflow where the library is pooled, edited and challenged before being resolved by FACS and next generation sequencing (12,13). However, library pooling is prone to bias as it selects for fast growth, and traits of interest might be lost if they are accompanied by a reduction in growth fitness. Similarly, when the best candidates are recovered from a one-pot screen, the top pool might be saturated by a few strongest variants, which limits the full network of interactions to be uncovered (14). A more useful output would be achieved if the members of the library could be assessed individually to produce the full overview of the genetic effects impacting the system.

Aside from one-pot approaches, screening can be achieved in a systematic manner by introducing a genetic modification in all strains of an existing yeast library (15,16); and to date, Synthetic Genetic Array (SGA) has been the state of the art method for such systematic analyses (17). SGA queries strains individually in a process based on arrayed mating, meiosis, sporulation and marker selection for the desired gene combination. This method has been particularly valuable for the identification of genetic fitness interactions among all double and a selected number of triple gene knock-outs (18,19). However, the reliance of SGA on meiosis is tedious, and the dependency of markers introduces experimental limitations. For example, sporulation is slow, 4–7 days, and the efficiency is difficult to control. Moreover, the high level of meiotic recombination required to ensure correct chromosome segregation in the first meiotic division (20) is per se undesirable. Firstly, as meiotic recombination mixes donor and recipient genomes, the genetic background of the donor and recipient strains need to be identical. Secondly, due to meiotic segregation, all genetic features need to be flanked by a genetic marker allowing selection of the desirable gene combination after meiosis. Such markers may influence expression of neighboring genes; and if many genes are studied, become a limitation for multiplexing. Thirdly, the high levels of meiotic Homologous Recombination (HR) may also set the stage for unwanted chromosome rearrangements if the strains suffer from repeated use of genetic elements, e.g. promoters. Lastly, flawed meiosis is known to frequently generate aneuploids (21,22), which may complicate further analyses. A faster method, independent of meiosis and marker selection, is therefore desirable.

We have developed a new screening platform that allows a genetic trait to be queried in large arrayed yeast libraries. Our method (Figure 1), CRI-SPA, combines Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology (23); (3) with selective ploidy ablation, SPA (24). In CRI-SPA, a (marker-free) genetic feature of interest is efficiently transferred from a single donor strain to the strains in a library in a process involving mating, Cas9 induced gene conversion, and haploidization by SPA. CRI-SPA can be massively parallelized using a pinning robot and executed in less than a week with as little as 4 hours handling time. Since it relies on image analysis for data extraction, the main cost associated with CRI-SPA comes from media and plastic consumables. Here, we use the method to transfer four genes responsible for the synthesis of the yellow plant metabolite betaxanthin, a biosensor for the morphine precursor L-DOPA (25) into the 4800 strains of the yeast knock-out (YKO) library (26). We demonstrate that multiple-gene transfer by CRI-SPA can be performed with and without selection in a highly reproducible manner. As a result, we unveil the full betaxanthin pathway-host gene interaction landscape and most prominently demonstrate that mutations impairing mitochondrial functions consistently influence betaxanthin levels. Finally, we show that CRI-SPA scores obtained in screening conditions correlate well with betaxanthin levels of reversed engineered hits in liquid medium.

MATERIALS AND METHODS

Strains and media

All strains constructed in this work are listed in Supplementary Table S1. Strains W8164-2B and W8164-2C of the SPA method (27) were kindly provided by Rodney Rothstein (Department of Genetics and Development, Columbia University, USA). The YKO library was acquired from Invitrogen. YPD and synthetic complete (SC) media, and SC drop out media were prepared as described by Sherman et al. (28), but with 60 mg/l L-leucine. For *URA3* counter selection, SC plates were supplemented with 1 g/l 5-fluoroorotic acid (5-FOA) and 30 mg/l uracil. To enhance the red phenotype of *ade2* cells, SC medium with only 4 mg/l adenine was used. To prevent *ade2* cells to be outcompeted by *ADE2* cells, growth media were supplemented with 40 mg/l adenine. Galactose and raffinose solutions were sterilized by filtration and used at 2% final concentration. For growth on solid medium, 20 g/l agar was added. For selection on solid medium, plates were supplemented with 200 mg/l geneticin (G418), 100 mg/l nourseothricin (NTC) and/or 200 mg/l hygromycin (HYG) as indicated. Half the concentration of the respective antibiotics were used in liquid cultivations. All strains were validated by diagnostic colony PCR to ensure correct integration at the intended chromosomal locus. For an experimental overview of CRI-SPA (CD) Donor strains construction, see Supplementary Figure S1.

Plasmid construction and PCR

Generally, DNA assembly was done by USER-fusion (29,30) employing uracil-specific excision reagent (USER™) enzyme from New England Biolabs. PCR fragments were generated using Phusion U Hot Start DNA Polymerase (Thermo Fisher). Details for construction of individual plasmids are provided in Supplementary Methods S1. All plasmids and primers used in this study are presented in Supplementary Table S2 and Supplementary Table S3, respectively. All plasmid maps are available for download as GenBank files.

Construction of yeast strains

All yeast transformations were carried out using the LiAc/SS Carrier DNA / PEG method described by Gietz and Schiestl (31). When antibiotic resistance genes were used as markers, cells were allowed to recover for 2 hours in liquid YPD prior to plating on solid selective media.

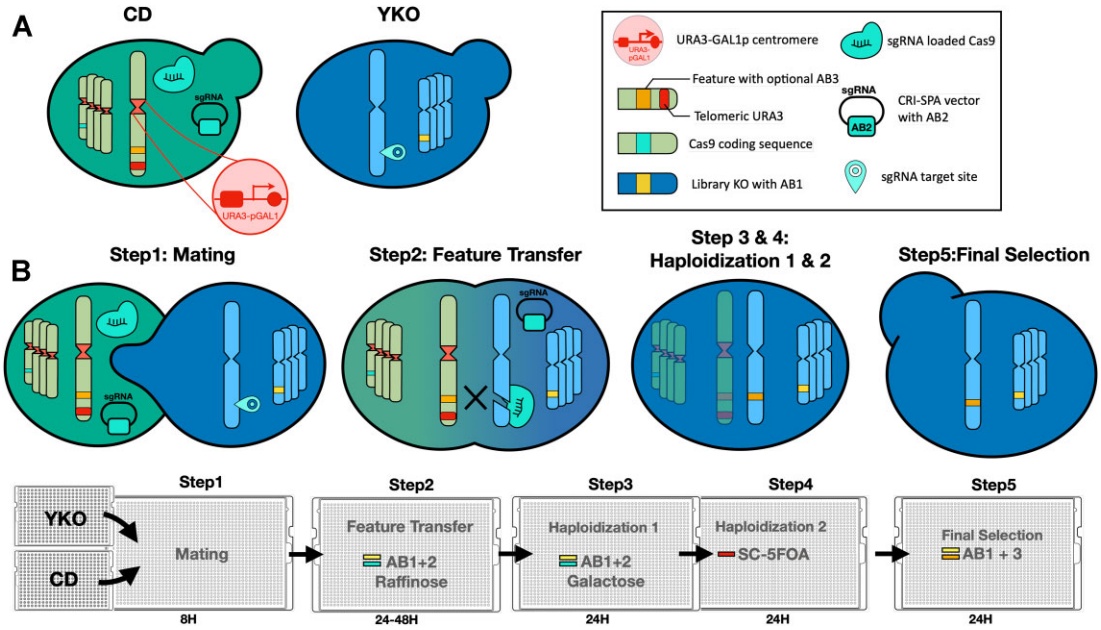


Figure 1. The CRI-SPA gene-transfer system. (**A**) Left: The CRI-SPA Donor (CD) strain and the recipient library strains (here YKO). Right: Graphics of the individual genetic components are shown in the legend box. The YKO strains contain a specific deletion marked by the antibiotic marker 1 (AB1) and a target site for the Cas9-sgRNA CRISPR nuclease. The CD strains contain: *Kl_URA3-GAL1p* cassettes at the centromeres of all chromosomes, a *cas9* gene, a selectable CRI-SPA plasmid maintained by antibiotic marker (AB2) and encoding a sgRNA targeting the insertion site of the genetic feature (blocked in the CD), a genetic feature of interest coupled to an antibiotic marker, AB3 (optional), and a *Kl_URA3* gene between the genetic feature of interest and the telomere. (**B**) The individual steps of the CRI-SPA procedure (see main text for details). **Step 1**, a CD strain is pinned onto all strains of the library plates and incubated for mating. **Step 2**, diploids are exhausted for glucose via growth on raffinose and selected for the marker in the library strains (AB1) and the marker on the CRI-SPA plasmid (AB2). In the diploid stage, the target site in the recipient strain is cleaved by Cas9-sgRNA. Repair of the resulting DNA DSB by gene conversion using the corresponding donor site in the CD strain as template transfers the genetic feature to the recipient locus. Note that the target site of the Cas9-sgRNA in the recipient chromosome is destroyed by insertion of the genetic feature of interest. **Step 3**, galactose induced SPA eliminates the donor chromosomes. **Step 4,** haploid cells containing only recipient chromosomes are selected on 5-FOA. **Step 5,** the modified library strains are obtained by selecting for AB1, which marks the YKO deletion, and optionally, by selecting for the genetic feature (AB3).

Universal CRI-SPA strains, UC strains, were made by inserting a gene-targeting substrate containing *cas9* and *LEU2* into chromosomal integration site X-3 of W8164-2B (*MATα*) and W8164-2C (*MAT***a**). The gene-targeting substrate was liberated from pHO8 by digestion with NotI (Thermo Scientific). The UC+:XII-5α strain (see Supplementary Figure S1 and Supplementary Table S1) was made from the UC strain (UC1α) by inserting an additional copy *of Kl_URA3* between expression site XII-5 and the telomere. Insertion was achieved via co-transforation with sgRNA plasmid pHO25 and a linear *Kl_URA3* repair substrate liberated from pCfB390 (32) using primers HOP181 and HOP182. CRI-SPA donor (CD) strain were made from UC strains. CD-*ade2*Δ was constructed by first transforming the UC strain (UC1α) with sgRNA plasmid pCfB2311 (33) and a linear repair substrate for disruption of *ADE2* via insertion of the *hphNT1* cassette, which was generated by PCR using plasmid pMEL12 and primers HOP64 and HOP65. The resulting strain was then further modified to harbor an additional *Kl_URA3* marker between the disrupted *ADE2* gene and the telomere by transformation with sgRNA plasmid pHO22 and a linear *Kl_URA3* repair substrate amplified from pCfB390 by primers HOP175 and HOP176, yielding CD-*ade2*Δ.

CD-Btx was constructed from UC+:XII-5α by transforming with the integrative plasmid pBTX1 harboring the betaxanthin-synthesizing genes CYP76AD1 and DOD, an NTC selectable marker, negative feedback resistant $ARO4^{K229L}$, $ARO7^{G141S}$ and an sgRNA expression cassette for targeting the XII-5 site. The BY-Btx strain was made by transforming the NotI excised insert of pBTX2, coding for CYP76AD1 and DOD and the negative feedback resistant $ARO4^{K229L}$, $ARO7^{G141S}$, into strain BY4741.

Introduction of CRI-SPA-hit knockouts into the BY-Btx strain was performed by replacing their Open Reading Frame (ORF) with a *kanMX* cassette (34). The *kanMX* cassette was amplified from pCfB2312 (32) and fused to targeting sequences up- and downstream of the relevant ORF. For the first round of KOs (DLD2, KGD2, MIP1, RCR2, YER084W, COX12, QCR10, SHY1, SGM1, VPS34, SSM4, UBC7, RPN4), the 400 bp up- and downstream targeting sequences were amplified using BY4741 genomic DNA as a template and by using primers with 40bp overhangs homologous to the *kanMX* cassette for amplification. The upstream, downstream, and *kanMX* amplicons were transformed into BY-Btx and fused *in vivo* by HR. For the second round of KOs (PRO2, VMA16, GET4, PRY1, STV1, YCR101C, YLR271W, GLR1, HEM25), the up and down homology sequences of the relevant ORFs were fused to the *kanMX* cassette by USER cloning. USER primers were used to generate PCR fragments of the up- and downstream sequences of the ORFs using genomic DNA from BY4741 as template. The relevant PCR fragments were USER-fused (29) to the *kanMX* cassette in *Escherichia*

coli and inserted into the USER cloning site of pCCM023, a modified version plasmid pCfB2909 lacking XII5-up and XII5-down homology sequences. The Up::*kanMX*::Dw gene-targeting substrates were released from purified plasmids by digestion with NotI prior to transformation into BY-Btx. All transformant BY-Btx strains were selected on solid YPD medium containing 200 mg/l G418. All KOs were verified by PCR amplifying the upstream and downstream integration sites of the *kanMX* cassette. For both amplifications, a primer annealing outside and a primer annealing inside of the cassette were used.

CRI-SPA high-throughput pin replication protocol

Automatic pin replication was carried out using the high throughput pinning robot ROTOR HDA from Singer Instruments (United Kingdom), along with replica pinning pads (RePads) and rectangular petri dishes (PlusPlates) from the same company.

For parallel mating of the CD Strain to the strains of the yeast deletion library, a single colony of the UDS was inoculated in 25 ml YPD supplemented with the appropriate antibiotic in order to maintain selective pressure for the gRNA plasmid. The UDS was grown overnight, washed by centrifugation, and resuspended with 25mL sterile water to remove the antibiotic. A volume of 150 μl of cell suspension was dispensed in wells of a 96-well plate serving as a source plate for the screen. From this source plate, the UDS was pinned in 1536 format on the mating plates. The arrayed strains of the deletion collection were pinned on top of the UDS. Plate shuffling was introduced at this stage to address experimental artifacts (see below). Mating was allowed to proceed from 8 h to overnight. Strains were then pin replicated onto a series of selective plates as outlined in Figure 1.

Image acquisition and processing

The data from our screen was extracted from images acquired with a specialized imaging system (Phenobooth, Singer) keeping acquisition parameters fixed (lighting power, camera brightness, gain, exposure, hue, saturation, white balances). To extract yellow color intensity, we applied a heuristic filter taking the geometric mean of Value and Saturation of image pixels in the HSV color space (Supplementary Figure S2A). This assumed that the colony's Hue was fixed (i.e. a colony can only harbour variants of yellow) which was our observation on YPD media.

We repurposed the mask functions of an image analysis package (35), to assign pixels to colonies on a plate image. The number of pixels assigned to a colony was used as a measure of colony size. Pixels belonging to a colony were filtered and the average filtered colony pixel value was used as the yellow intensity score.

Randomisation of positional artifacts, outlier removal, data normalization

We observed that colony size and yellowness score were subject to positional effects known to affect colony size in high density array screens (36). We adopted a colony position randomization strategy to randomize plate, agar quality and neighbor effects. This randomization was done when pinning the YKO library from its storage 384 format to the screen's 1536 quadruplicate format. Each one of the four 384 arrays from a given library plate was pinned on a different screening plate (Supplementary Figure S2B). As a result, each colony quadruplicate was positioned on a different screen plate and exposed to randomized agar qualities and neighbor effects. Still, row-column colony positions were not altered by this randomization, and colonies were therefore still affected by non-random edge effects. These were corrected by setting the median of colonies in outer frames equal to that of center colonies (37). After edge correction, the screen's data was centered by subtracting the screen's mean from each data point and normalized by dividing all data points by the standard deviation of the screen. After this normalization, outliers within quadruplicates were detected with Grubbs' test for outliers and deleted. Genes represented by two colonies or less were deleted from the analysis. Finally, the mean of the colony replicate was used as the final score for each gene.

Combining the data of multiple screens

To combine the data of different screens, the edge-corrected data for each individual screen was normalized by subtracting the screen mean and by dividing colony scores by the screen standard deviation. At this point, the individual datasets were pooled and outliers removed with Grubbs' test. After outlier removal, mean yellow intensity and size of gene-deletion strains with three or more colonies was used as the final yield and fitness, respectively, (pipeline and data can be accessed on the repository https://github.com/pc2912/CRI-SPA_repo).

Betaxanthin fluorescence quantification

Fluorescence in SC medium was measured by a Synergy Mx plate reader (Agilent BioTek) using excitation of 465/20 nm and emission of 525/20 nm as described previously by Savitskaya *et al.* (14).

Gene enrichment analysis and GO term graphs

Gene enrichment analysis was performed using GOATOOLS (38,39) to generate the gene groups indicated in Figure 5. The default Benjamini-Hochberg correction was used to account for the number of tests, and a P-value of 0.05 was used as a threshold to accept GO terms as enriched. For visualizing results at a systems-level, a graph was constructed where nodes representing GO terms were connected if they shared associated genes. The size of the nodes was drawn as a function of the significance of the term enrichment and the edge opacity was drawn in relation to the number of genes shared between terms. Nodes were colored according to their description or the description of their ancestors (i.e. descriptions contained key words such as 'mitochondria', 'translation', ect.). Finally, the resulting networks were built in networkX, transferred to Cytoscape (40) with py4cytoscape and minor manual

adjustments were made for visibility (scripts available at https://github.com/pc2912/CRI-SPA_repo).

Statistical tests

Statistically significant overlaps of hits identified in screen A and screen B were determined by using a test based on hypergeometric distribution: p_h($x \geq k \mid M,n,N$). M is the number of genes quantified in both screen A and screen B, and $n = 192$ is the number of genes defined as hits in Screen A. $N = 192$ is the number of genes defined as hits in screenB. Finally, k is the number of overlapping gene hits between the two screens. Statistically significant increases (or decreases) in yellowness and fluorescence between BY-Btx and BY-Ref were determined by a one-sided t-test. Significant differences in yellowness score on solid YPD medium and fluorescence in SC medium between BY-Btx and reversed engineered hit mutants were identified with a one-sided Welsh-tests.

RESULTS

Basic components of the CRI-SPA system

CRI-SPA is based on four components: (i) Universal CRI-SPA strains (UC strains), see Supplementary Table S1, which are SPA strains (24) equipped with a *cas9::LEU2* cassette, see Materials and Methods, in integration site X-3 (41). A SPA strain contains a *Kl_URA3-GAL1*p cassette next to all of its sixteen centromeres; and all chromosomes of a SPA strain can therefore be destabilized on media containing galactose. Hence, in diploids where a SPA strain serves as one of the parent strains, SPA chromosomes can be selectively eliminated on galactose medium followed by counterselection on 5-FOA (24). As a result, a haploid strain composed by chromosomes solely originating from the non-SPA parent strain is achieved. We have constructed UC strains of both mating types. (ii) CRI-SPA vectors, see Supplementary Table S2, which are selectable 2μ based plasmids encoding a desirable gRNA. CRI-SPA vectors enable formation of Cas9–gRNA complexes that serve two purposes. Firstly, they mediate integration of a genetic feature of interest into a specific site in a UC strain; and secondly, they enable HR mediated transfer of the genetic feature during CRI-SPA (see below). In addition, the marker gene on the CRI-SPA vector facilitates selection for diploid cells during the CRI-SPA procedure, see below. (iii) CRI-SPA Donor strains (CD), see Supplementary,Table S1, which are UC strains that contain the genetic feature of interest in a defined chromosomal locus, a CRI-SPA vector encoding a gRNA matching the wild-type sequence of the modified locus, and a *Kl_URA3* gene between the genetic feature of interest and the telomere of the modified chromosome. This additional *Kl_URA3* gene is used to counterselect undesired recombination events that may accompany Cas9 induced gene conversion, see below and Supplementary Figure S3. (iv) A CRI-SPA compatible recipient strain library. The strains in the library need to contain a genetic marker allowing selection for diploid cells when they are mated to the CD strain. Moreover, in the current version of our CRI-SPA system, the library needs to be composed of strains that are *ura3* as SPA depends on counterselection of *Kl_URA3*. In the present study, we have used the genome-wide YKO collection as the recipient library (26). The strains in this library are all *ura3Δ0* and the individual gene deletions are marked by *kanMX* (42), which can be selected by addition of G418 to the medium.

Experimental steps of the CRI-SPA method

The CRI-SPA procedure is performed in five steps during six days (Figure 1). In step 1, The CD strain and the recipient library strains mate overnight on YPD medium to form diploid cells. In step 2, diploid cells are selected via a complementary marker setup for 48 h on solid raffinose. In the present study, this was achieved by replica-pinning the colonies from YPD to solid raffinose medium containing G418 and either nourseothricin (NTC) or hygromycin (HYG) to select for *kanMX* marked deletions delivered by the parental library strains and the *natMX*- or *hphNT1*-based CRI-SPA plasmid delivered by parental CD strains, respectively. In diploid cells, the Cas9-sgRNA CRISPR nuclease produces a DNA DSB at the recipient locus in the library strain chromosome, and this DNA DSB is repaired through HR using the corresponding modified locus of the CD strain as a repair template. Most commonly, repair is expected to proceed via synthesis-dependent strand-annealing (43) resulting in desirable transfer of the genetic feature from the donor to the recipient chromosome by gene conversion. However, we note that undesirable chimeric donor/recipient chromosomes may be produced as the result of repair by gene conversion accompanied by a crossover or if repair is mediated by break induced replication (44). In CRI-SPA, these undesirable repair outcomes are counter-selected in step 4 (also see Supplementary Figure S3). The suboptimal and non-repressing carbon source raffinose (45) was used to exhaust glucose and to provide a slow growth step thereby allowing more time for CRISPR mediated gene transfer compared to growth on glucose. In step 3, the absence of glucose repression allows for the sudden induction of the *GAL1* promoter as the colonies are replica-pinned from raffinose to solid galactose medium (45). Active *GAL1* promoters disrupt the centromeres of all CD strain chromosomes, hence, inducing CD strain chromosome loss as the cells are dividing (24). In step 4, cells that have lost all donor chromosomes are selected by transferring the strains to solid SC medium containing 5-FOA. Note, this medium also counterselects undesired recombinant strains resulting from DNA DSB repair involving crossing-over or break-induced replication at the target locus (Supplementary Figure S3). As a result, step 4 generates haploid strains that solely contain the chromosomes originating from the recipient strain as well as the genetic feature of interest. In the final step 5, recipient cells are selected by the marker of the library, in this case by the *kanMX* marker of the YKO library to eliminate CD cells (if any) that did not mate, and which survived *kanMX* and 5-FOA selection. Optionally, the genetic feature of interest may also be selected for if it is accompanied by a marker. This may eliminate unmodified recipient cells (if any) that have escaped the selection step for diploids. Conveniently, we note that CRI-SPA can be executed by hand or scaled with a pinning robot to

accommodate a range of budgets and throughputs. A more detailed scheme describing the five CRI-SPA steps is provided in Supplementary Methods S2.

CRI-SPA donor strain construction and test for cas9 induced DNA DSB formation efficiency

CD strains are constructed in three simple steps, see Supplementary Figure S1. Firstly, an additional *Kl_URA3* marker is inserted into a locus between the integration site of the genetic feature of interest and the telomere. This is achieved in a CRISPR mediated process based on a gRNA plasmid expressing the appropriate gRNA. Secondly, the resulting strain is cured of the gRNA plasmid. Finally, the CD strain is obtained by inserting the genetic feature into the desired locus in a second CRISPR mediated reaction assisted by the CRI-SPA vector.

Successful CRI-SPA mediated allele transfer depends on highly efficient DNA DSB induction at the target locus by a CRISPR nuclease directed by a specific gRNA encoded from the CRI-SPA vector. Hence, prior to a CRI-SPA experiment, it is important to demonstrate the efficiency of a desired CRISPR nuclease/gRNA complex. This can be judged by performing a TAPE (Technique to Assess Protospacer Efficiency) experiment (46), which examines lethality of a UC strain after transformation with a CRI-SPA plasmid in the absence or presence of an efficient repair template. Compared to a reference plasmid, a CRI-SPA vector suitable for a CRI-SPA experiment needs to produce transformants at numbers that are several fold higher in the presence than in the absence of a repair template, see Figure S4. Note that the TAPE test can conveniently be combined with CD strain construction by using a repair template that is designed to introduce the genetic feature of interest into the desired locus of the UC strain.

Transfer of ade2δ::hphNT1 into a set of arrayed gene-deletion mutants

To examine the potential of CRI-SPA, we first investigated whether it is possible to transfer a HYG selectable *ade2*Δ*::hphNT1* cassette from a CD-*ade2*Δ strain to a subset of mutants from the YKO library containing G418 selectable gene-deletion cassettes. Advantageously, *ade2*Δ strains produce an easy-to-score red phenotype (47,48), which may be epistatically blocked by mutations upstream in the purine pathway, e.g. *ade3* (49). Successful interchromosomal transfer of *ade2*Δ*::hphNT1* into YKO library mutants should therefore produce mostly red colonies, but should also be able to identify epistatic genetic interactions, like *ade3*Δ, which should appear as white colonies. Hence, we set out to investigate whether CRI-SPA would be able to quickly identify mutants that are epistatic to *ade2* by screening for double mutants that form white colonies.

The CD-*ade2*Δ strain was constructed as part of a TAPE test demonstrating that the NTC selectable *ADE2* CRI-SPA vector (pHO24) is able to produce a gRNA that allows Cas9 to efficiently cleave the *ADE2* locus (Supplementary Figure S4). Specifically, we obtained >6-fold more transformants in the presence (915 transformants) than in the absence (145 transformants) of the *ade2*Δ*::hphNT1* cassette repair template. Next, the resulting CD-*ade2*Δ strain was used in a CRI-SPA experiment to screen the strains on plate 9 of the YKO library (*MAT*$\underline{a}$). This plate contains 376 gene deletions, including the negative interaction control *ade3*Δ*::kanMX*. The *ade2*Δ*::hphNT1* transfer experiment was performed as a 2 × 2 quadruplicate using 1536 format. In all growth steps, except the final one, media was supplemented with additional adenine to reduce the negative fitness effects due to the *ade2* deletion (50). On the final plate, the medium contained reduced adenine levels to speed-up development of red color, see MATERIALS AND METHODS.

After the CRI-SPA experiment, virtually all quadruplicates were red indicating that the *ade2*Δ*::hphNT1* cassette was efficiently transferred to the vast majority of all mutant strains of the library, see Figure 2. Only a few mutant strains did not produce any growth after CRI-SPA and such mutants may represent strains that failed to go through CRI-SPA, e.g. if they failed to mate. For a few quadruplicates, some of the colonies appeared white/pinkish. Further restreaking on solid selective media (G418 and HYG) demonstrated that they contained a mix of cells that were able to form white and red cells. Moreover, cells forming white colonies were able to grow in the absence of adenine and diagnostic PCR demonstrated that their genotype was *ADE2* and *ade2*Δ*::hphNT1*. Further, full genome sequence analysis of selected purified white strains showed that they were aneuploid or diploid strains containing two copies of the target chromosome of which only one has received a copy of the *ade2*Δ*::hphNT1* cassette, see Supplementary Figure S5. Formation of such cells may reflect that many strains in the gene deletion library suffer from aneuploidy (51). Some mutants may be particularly prone to this phenomenon and the most prominent case of a 'false whitish' phenotype was observed for *scp160*Δ, which is known to display ploidy instability (47,52–53). Alternatively, they may be derived from diploid cells resulting from endoreduplication of the recipient genome, which may happen during SPA (27), see Supplementary Figure S6. Although formation of *ADE2*/*ade2*Δ*::hphNT1* cells potentially could complicate the CRI-SPA analysis as they formed white cells, in particular as they display a fitness advantage over *ade2*Δ*::hphNT1* cells, this was generally not the case. Indeed, the *ade2*Δ*::hphNT1 ade3*Δ*::kanMX* quadruplet was easily identified as the only example of a quadruplet where all four colonies were purely white, see Figure 2. Hence, we believe that formation of this cell type is quite rare and they may likely only impact experiments if they display a significant fitness advantage. Altogether, we conclude that that the CRI-SPA method was able to efficiently transfer the *ade2*Δ*::hphNT1* cassette from a CD to a recipient library and that it was able to robustly identify the negative *ade2*Δ*::hphNT1 ade3*Δ*::kanMX* interaction.

Establishment of an efficient CRI-SPA based betaxanthin screening platform

We next sought to demonstrate the usefulness of the CRI-SPA technology in a metabolic engineering context. We attempted to map out the genetic requirements, at single gene resolution, for the production of betaxanthin,

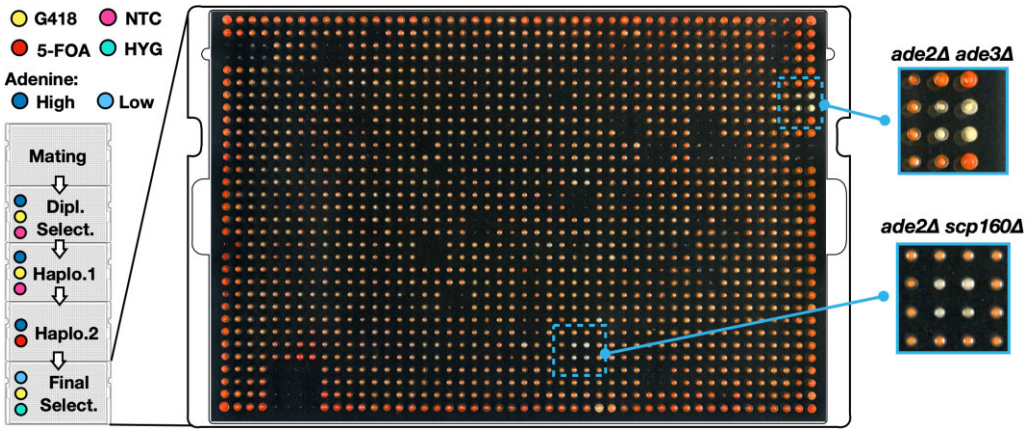


Figure 2. Transfer of the *ade2Δ::hphNT1* cassette from a CD to plate 9 of the YKO library. CRI-SPA mediated transfer of the *ade2Δ::hphNT1* cassette to the mutant strains of plate 9 of the YKO library was performed in quadruplicate. The final CRI-SPA plate was photographed and shown in the middle. Selection applied at the different CRI-SPA steps are indicated to the left by the color code. G418 selects for the *kanMX* marker in the gene deletion cassettes of the YKO library, NTC selects for the CRI-SPA plasmid, HYG selects for the *ADE2* disruption cassette *ade2Δ::hphNT1*, and 5-FOA counter-select donor strain chromosomes, which all contain *Kl_URA3* markers. The four *ade2Δ::hphNT1 scp160Δ::kanMX* and the four *ade2Δ::hphNT1 ade3Δ::kanMX* CRI-SPA colonies are framed by blue squares. Magnifications of the *ade2Δ::hphNT1 ade3Δ::kanMX* quadruplets are shown to the right as indicated.

a plant metabolite derived from the shikimate pathway. Betaxanthin is yellow and fluorescent; and has previously been used as a biosensor and a proxy for production of the morphine precursor L-DOPA in yeast (25). To test whether betaxanthin production could be meaningfully evaluated in a CRI-SPA screen using the YKO library, we firstly introduced a five-gene betaxanthin-cassette (Btx-cassette) into BY4741, producing strain BY-Btx. The Btx-cassette harbors *CYP76AD1* and *DOD* (25), which are required for production of betaxanthin, the dominant mutant genes $ARO4^{K229L}$, $ARO7^{G141S}$, which alleviate the negative feedback regulation of the shikimate pathway (25); (54), as well as the selectable *natMX* marker. As expected for a strain producing betaxanthin, BY-Btx, unlike the reference strain BY-Ref (BY4741), produced yellow colonies (Figure 3A); and liquid cultures were yellow and fluorescent (Figure 3B). Importantly, betaxanthin production did not appear to influence fitness, as BY-Btx displayed essentially the same growth rate as BY-Ref Supplementary Figure S7.

We next tested whether significant differences in betaxanthin production in the YKO library background could be scored in a simple colony-based assay suiting the format and throughput of a CRI-SPA experiment. In our hands, the colony color phenotype was the easiest to score with a color filter, extracting color intensity from RGB pixel values, see 'Image acquisition and processing' in Materials and Methods and Supplementary Figure. S2A). Using this method, we demonstrated that the yellowness-score of colonies obtained with the betaxanthin producing strain was around 5-fold higher ($P \leq$ 10e–10) than the score obtained with colonies of the reference strain, see Figure 3A. We also examined production of betaxanthin in liquid SC-medium by measuring betaxanthin fluorescence in a plate-reader assay using a method previously described by Savitskaya *et al.* (14). In this setup, the BY-Btx strain was 6-fold ($P \leq 0.007$) more fluorescent than the reference strain, see Figure 3B, demonstrating that the method could be used as a complementary assay for characterizing specific Btx-hit strains identified by yellow color development on solid medium. In summary, we have established a readily scalable image analysis method able to quantify betaxanthin production based on yellow color intensity; and we note that yellow color and fluorescence are widely accepted betaxanthin detection methods (14,25,55).

To set the stage for a genome-scale betaxanthin CRI-SPA experiment, we firstly constructed a betaxanthin CD (CD-Btx) that harbors the NTC selectable five-gene Btx-cassette in the XII-5a expression site, a *hphNT1* selectable CRI-SPA plasmid (pHO29), which encodes a gRNA that efficiently targets empty XII-5a sites (56), and a *Kl_URA3* marker positioned between XII-5a and the telomere. Secondly, given the considerable size, 9.3kb, of the Btx-cassette we performed a pilot experiment to assess how efficient it is transferred from the CD-Btx strain to the mutant strains of plate 9 of the YKO library, which we also used in the *ade2::hphNT1* transfer experiment described above. Lastly, as the Btx-cassette, unlike *ade2Δ::hphNT1*, does not confer a significant fitness defect in BY-Btx the YKO reference strain, we decided to explore the possibility of transferring the Btx-cassette without selecting for its *natMX* marker. Hence, we performed the pilot CRI-SPA Btx-cassette transfer experiment in quadruplicate in the absence of nourseothricin. Out of the 376 strains on the library plate, 373 strains (>99%) produced colonies after CRI-SPA, and 367 (>97%) produced all four colonies, see Supplementary Figure S8. Encouragingly, the vast majority of colonies were yellow indicating highly efficient transfer of the Btx-cassette despite that we did not apply nourseothricin for its selection during any of the steps in the CRI-SPA procedure.

CRI-SPA colonies are not clones, and strain-to-strain betaxanthin production comparisons can therefore be compromised if some colonies are contaminated with significant amounts of other cell types producing levels of betaxanthin that are different as compared to the levels of the modified recipient mutant cells. Since we did not select for the Btx-cassette in the final CRI-SPA step, we rea-

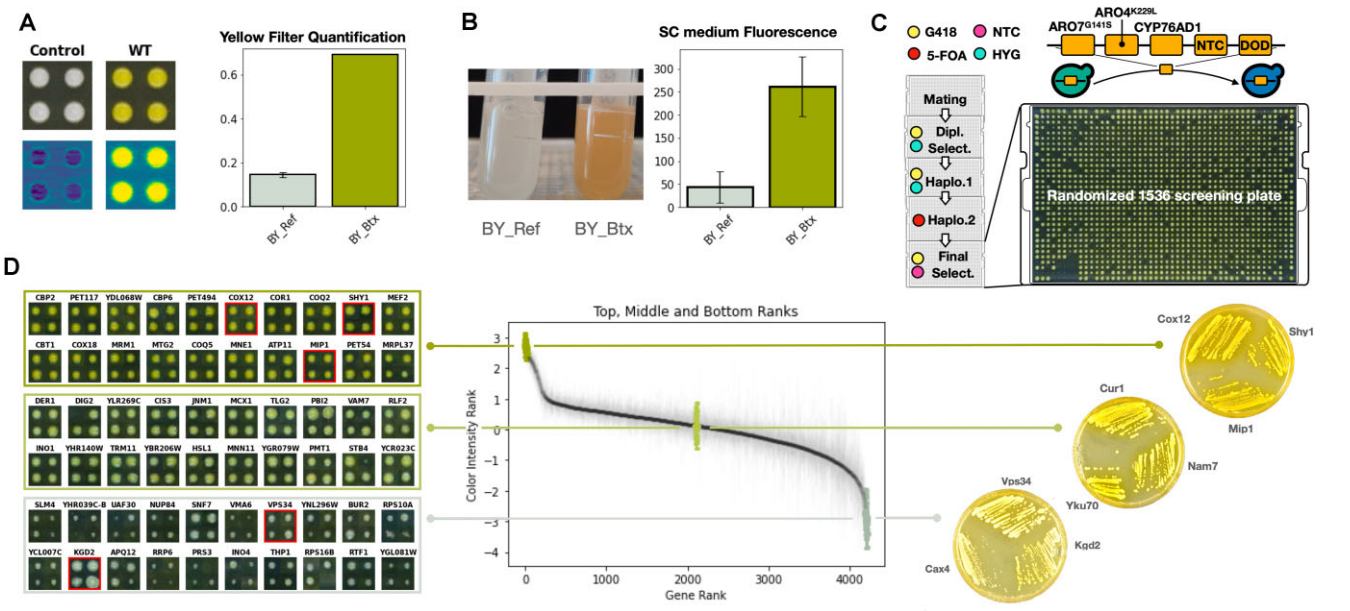


Figure 3. Betaxanthin production in YKO strains. (**A**) Left: images of BY-Ref and BY-Btx colonies in visible light (top) and corresponding filtered images (bottom). Right: quantification of yellowness of BY-Ref and BY-Btx colonies by image analysis, see main text and Supplementary Figure S2 for details. (**B**) Left: image of BY-Ref and the BY-Btx strains in liquid SC medium. Right: quantification of fluorescence produced by BY-Ref and the BY-Btx strains in liquid SC medium. (**C**) Final plate of a CRI-SPA Btx-cassette transfer experiment. At the top, a scheme illustrating the Btx-cassette and how it is transferred from the CD strain to a library recipient strain. To the left, diagrams with color codes indicating selection applied at the different CRI-SPA steps. G418 selects for the *kanMX* marker in the gene-deletion cassettes of the YKO library, HYG selects for the CRI-SPA plasmid, NTC selects for the Btx-cassette, and 5-FOA counter-select donor strain chromosomes, which all contain *Kl_URA3* markers. (**D**) In the middle, yellow intensity ranking of all YKO mutant strains after CRI-SPA mediated Btx-cassette transfer. To the left, mosaic images of the four individual replicates obtained with selected YKO mutants displaying higher, similar, or lower yellowness as compared to BY-Btx. Mutants in red frames were purified and restreaked on solid YPD medium, see images to the right.

soned that the largest source of error would be contributed by white unmodified recipient cells, which were not eliminated during CRI-SPA steps 2–3, and which would reduce the yellowness score of the mutant colony. To explore the extent of this potential problem, we picked cells from each of the four colonies obtained for each strain on the final CRI-SPA screen plate (YPD-G418) and replica spotted approximately 1000 (two of the colonies) and 10 000 cells (the other two colonies) on YPD + G418 (control) and YPD + G418 + NTC plates, see Supplementary Figure S8. Contaminating unmodified YKO cells should form white colonies on YPD + G418, but not on YPD + G418 + NTC medium. For 360 of the strains (96%), sufficient viable cells were transferred (> 100 colonies produced in three of the four trials) to allow for further analysis. Next, we qualitatively compared the number of colonies formed on YPD + G418 and YPD + G418 + NTC plates for all trials. This analysis identified eight strains (∼2% of the 360 strains), which in one or more trials contained >1% unmodified YKO library cells in the colonies of the final CRI-SPA screening plate. Most of these colonies contained little if any modified YKO cells and these trials therefore represent rare flawed CRI-SPA transfer trials. We note that for six of these eight strains ≥ two of the four trials were not contaminated with unmodified YKO cells. Accordingly, we conclude that transfer of unmodified YKO cells through the CRI-SPA procedure is a minor problem.

We also reasoned that yellow CD-Btx and diploid cells, which survived CRI-SPA steps 4–5, would compromise strain-to-strfain betaxanthin production comparisons if they were present in large amounts in the CRI-SPA colonies. We therefore examined the final CRI-SPA plates for colonies that contained CD-Btx and were diploid. In parallel with the experiment presented above, cells were therefore spotted in the same way, but on solid SC-Ura medium, which allows CD-Btx and diploid cells to grow, and on solid SC-Ura + G418 medium, which allows diploid, but not CD-Btx, cells to grow. These analyses demonstrated that basically no CD-Btx cells survive the CRI-SPA screen. Some diploid cells survive, but they constitute less than 0.1% of the cells in the colonies. For a few strains, one of the four trials contained ∼1% diploid cells. Hence, we conclude that carryover of CD-Btx and diploid cells to the final CRI-SPA plate constitutes a minor problem in this CRI-SPA experiment. Lastly, we note that the *scp160*Δ strain, which produced a false phenotype in the *ade2*Δ*::hphNT1* cassette transfer experiment failed to produce more than a few surviving cells after CRI-SPA in this experiment.

Introduction of the betaxanthin pathway into a genome-wide library of gene-deletion strains

With the CRI-SPA based betaxanthin-pathway transfer system and a betaxanthin quantification method in place, we then used CD-Btx to transfer the Btx-cassette to the 4787 unique deletion strains of the YKO library using CRI-SPA in quadruplicate amounting to a total volume of 19148 transfer events. To address positional effects known to affect colony size in high density array screens (36,37), we randomized the YKO library arrays (in 384 format) when

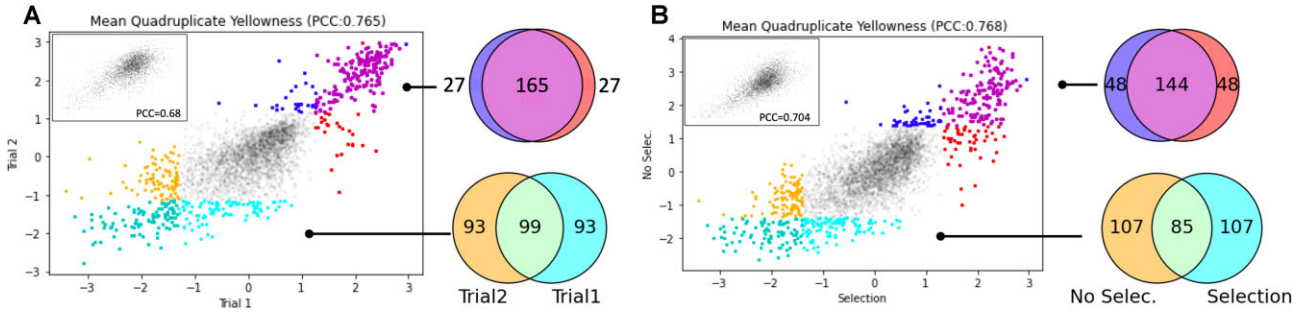


Figure 4. Reproducibility of betaxanthin CRI-SPA experiments in the presence and absence of pathway transfer selection. (**A**) The mean yellow color intensities obtained for each mutant in repeat 1 and repeat 2, which both were performed in the presence of *natMX* to select for the Btx-cassette genes. (**B**) The mean yellow color intensities obtained for each mutant in trial 1 and trial 3. Trial 3 was performed in the absence of *natMX*. The top and bottom 192 genes and their overlap across screens are colored and displayed in flanking Venn diagrams. Insets show the correlation for colony fitness across screens.

scaling up to 1536 quadruplicate format. Specifically, each of the four 384 arrays from a given library plate were pinned on a different screening plate (see Supplementary Figure S2B and Materials and Methods for full description of the image analysis, colony randomization and data processing pipeline).

In the first set of experiments, we employed NTC selection in the final step of CRI-SPA as a safety measure to ensure that all cells contain the Btx-cassette. Encouragingly, the vast majority of the colonies appeared yellow after CRI-SPA indicating that the CRI-SPA mediated transfer of the betaxanthin pathway into the library was successful (Figure 3C). We extracted the yellowness score of all colonies and after removing remaining positional artifacts and outliers (see MATERIALS AND METHODS including accompanying python scripts available on our CRI-SPA repository), 4224 strains ($\approx$ 88% of the library) produced $\geq$ three viable CRI-SPA colonies and were henceforth included in our analysis.

We finally ranked the mean yellowness for all genes (Figure 3D) and found that the screen produced a strong yellowness signal to noise ratio (the *Z*-scores of the 16 top and bottom hits was 40.83, –6.93; respectively.). This analysis demonstrated that our screen can produce strains with a large range of betaxanthin synthesis abilities.

Since CRI-SPA colonies are not monoclonal, we isolated three candidates representing three different color levels and restreaked them on solid YPD medium with appropriate selection. As expected, the individual streaks of cells formed colonies that were uniformly yellow indicating little, if any, presence of contaminating cell types. Importantly, the color differences between the sets were easily distinguished on plate images, validating the position of the mutant strains in the color ranking scheme (Figure 3D).

Screen to screen reproducibility and marker-free CRI-SPA

Confident that the screen readout was accurate for a handful of strains, we repeated the entire CRI-SPA experiment to validate its reproducibility. After CRI-SPA, the yellowness of the individual mutants obtained in this experiment, trial 2, were then plotted against the yellowness obtained for the mutants in the first experiment, trial 1. The clear linear correlation between the two trials (PCC = 0.765) demonstrates the reproducibility of CRI-SPA (Figure 4A). Moreover, we found that, from all successfully labeled genes in the two screens, 86% (P = 1.09e–220) and 52% (P = 9.97e–83) of the 192-top and 192-bottom hits were common in the two trials, respectively. Hence, and especially for the top-hits, reproducibility was high.

In a third experiment, trial 3, we explored the possibility of transferring the five-gene Btx-cassette from CD-Btx to the full YKO library without selecting for the *natMX* marker, which is present in the Btx-cassette. In this screen, the CRI-SPA plasmid's restriction of the target site maintained the sole selection for genetic transfer. With this setup, 95.1% of mutants produced at least three colonies after CRI-SPA mediated transfer of the Btx-cassette. Importantly, the yellowness score in this trial had a clear linear correlation with trial 1 (PCC = 0.768, Figure 4B). Moreover, when the 192 most yellow and 192 least yellow mutants obtained in the two trials were compared, the top and bottom groups had 75% ($P \leq$ 10e–180) and 44% ($P \leq$ 10e–69) overlap across screens, respectively (Figure 4B). Hence, the above analyses demonstrate that CRI-SPA is reproducible and that the combined effect of a mutation and a genetic feature can be reproduced even without selecting for this feature.

Systems perspective of betaxanthin production

In contrast to one-pot screens, CRI-SPA produces a readout for each individual strain within a library. We investigated whether this systematic gene labeling could reveal new system level patterns in betaxanthin production. For this analysis, we combined the data of four screens (see supplementary part), obtaining a CRI-SPA score for 4761 genes out of the 4788 present in our library with a median of 15 replicates (16 is maximum) per gene. We first plotted yellowness of all strains against colony size using the latter feature as a proxy for fitness, see Figure 5. Plotted in this way, the vast majority of cells map in the close vicinity of the WT, indicating that the mutations in these strains do not impact betaxanthin production (Figure 5). We next grouped the remaining mutant strains according to their yellowness score: the 'yellow' group was defined as mutants with the highest level of betaxanthin production (1.2 standard deviation greater than the screen mean, yellow points in Figure 5). The 'white' group was defined as mutants with the lowest

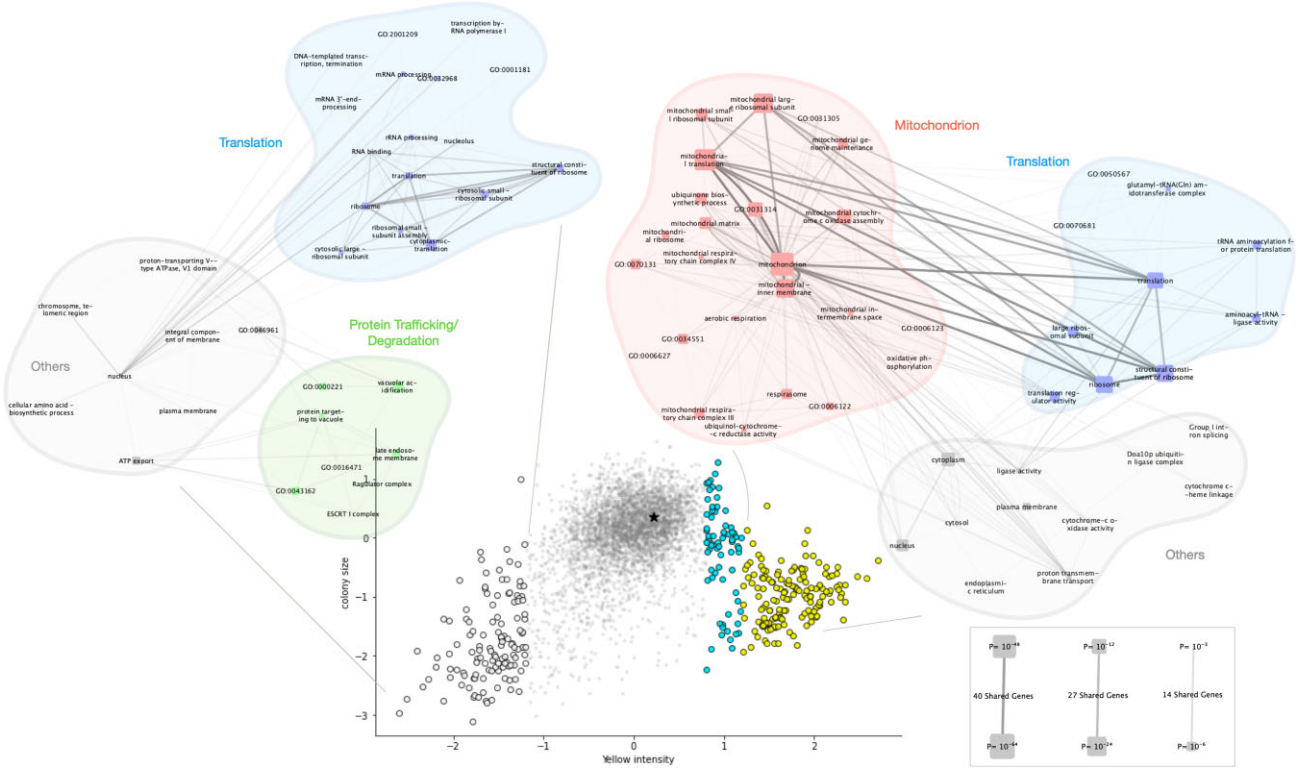


Figure 5 Betaxanthin yield and fitness relationship reveals mechanistic patterns in betaxanthin production. Colony size (i.e. fitness) vs Yellow Color intensity (i.e. Betaxanthin Yield). The results of Gene Ontology Enrichment Analyses on several subgroups of genes are shown as graphs for the bottom hits (white, 1.2 std below the screen mean) and top hits (yellow, 1.2 std above the screen mean). As indicated by the bottom right scale, node size indicates the significance of GO terms enrichment, edge transparency indicates the number of shared genes between GO terms. The cyan group did not show any term enrichment. The position of the BY-Btx strain is marked with a black asterisk.

level of betaxanthin production (1.2 standard deviation below the screen mean, white points in Figure 5). We observed that most members of the yellow group displayed a strong fitness penalty. We therefore defined a third group of mutants, the 'cyan group', see Figure 5. This group contains mostly highly yellow strains with little or no fitness penalty. Next, we examined whether specific cellular functions could be linked to betaxanthin levels by running gene ontology enrichment analyses (38) within the three groups. The cyan group did not produce any enrichment. This may not be surprising as the hits in this group were the least reproducible (Supplementary Figure S9). The white group, despite its broad fitness range, was enriched for terms describing the degradation machinery, translation, and vacuolar mechanisms. Most strikingly, the yellow group displayed a dramatic enrichment for genes associated with mitochondrial functions, particularly the respiratory chain, which might explain the fitness penalty observed in this group. Hence, this analysis demonstrates that CRI-SPA is able to identify a novel network of genes, which impact functions of a specific organelle in the cell, and which, when deleted, lead to reduced fitness and high betaxanthin levels.

Validation of CRI-SPA hits in liquid culture

Currently, CRI-SPA screens for phenotypic changes on solid medium and it is possible that a phenotypic hit may potentially vanish in different growth conditions. To explore that possibility, we decided to test how selected Btx-hit strains fare in liquid SC-medium. Moreover, to further validate that strains picked up in a CRI-SPA screen truly reflect the impact of the specific mutation provided by the YKO library strain, we re-introduced the mutations in the BY-Btx reference strain by conventional one-step gene-targeting (57) using *kanMX* to select for the deletion. Hence, we reintroduced 16 top and 6 bottom KO hits in the reference strain BY-Btx, covering a variety of cellular functions and strain fitnesses (Figure 6A, Supplementary Table S4). To avoid penalizing slow growing strains, and strains with a prolonged lag-phase, we grew the reference and the 22 mutant BY-Btx strains in liquid SC medium with 2% glucose for 7 days. This setup is possible as betaxanthin does not appear to be catabolized during fermentation as its levels remained stable throughout the stationary phase of a BY-Btx pilot experiment (Supplementary Figure S7).

We then examined all strains for betaxanthin production using a fluorescence readout at the fermentation end-point. The results showed that betaxanthin levels strongly correlated with the yellowness score obtained for the same mutants on the final CRI-SPA selection plate (PCC = 0.680, see Figure 6B). For the positive hits, of the 13 scoring significantly higher than the WT for yellowness in the CRI-SPA screen, nine were significantly more fluorescent than the reference BY-Btx strain, indicating that they also produced

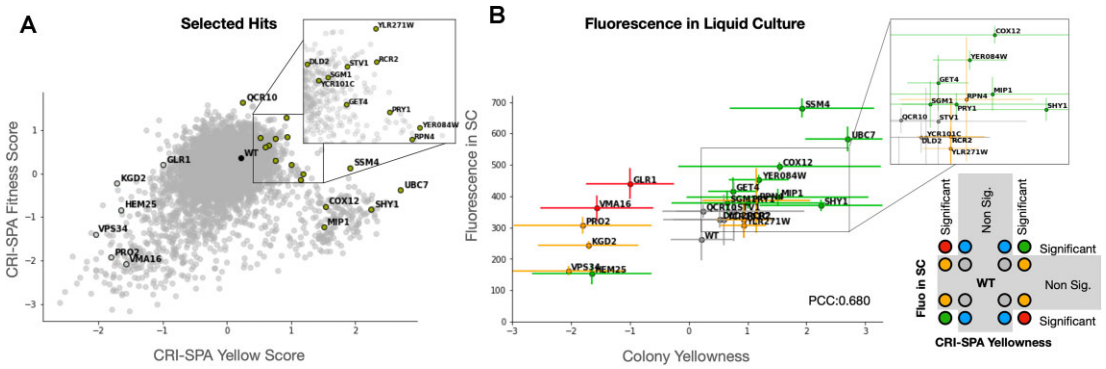


Figure 6. CRI-SPA hits identified on solid medium produce enhanced amounts of betaxanthin in liquid medium. (**A**) selected Btx-hit strains yellow and fitness scores within the screen. (**B**) Correlation between CRI-SPA yellowness score and fluorescence of reversed engineered hits after 7 days cultivation in liquid SC medium. Error bars show standard deviations for 16 and 3 replicates for CRI-SPA and liquid Fluorescence, respectively. Markers coloring scheme, below inset, is defined by the agreement between CRI-SPA score and fluorescence in liquid medium, and significance difference from WT reported by a one-sided Welch's test.

significantly increased levels of betaxanthin in liquid SC medium. Indeed, the two most yellow mutant strains identified in the CRI-SPA screen, *ubc7*Δ and *ssm4*Δ/*doa10*Δ, were 2.6-fold and 2.2-fold more fluorescent than BY-Btx (*ssm4*Δ/*doa10*Δ: $P \leq 6 \times 10e{-}4$, *ubc7*Δ: $P \leq 2 \times 10e{-}3$), respectively. Another set of positive CRI-SPA Betaxanthin hits were composed by *mip1*Δ, shy*1*Δ, *YER084W*Δ, and *cox12*Δ. The corresponding genes all encode mitochondrial proteins showing that dysfunctional mitochondria may also stimulate betaxanthin production in liquid SC medium. For the negative hits, the *hem25*Δ hit was the only strain displaying significantly reduced fluorescence compared to BY-Btx (0.59-fold, $P \leq 0.05$). Taken together, we show that CRI-SPA is able to identify gene hits, which perform in growth conditions routinely used for metabolic engineering strain development.

DISCUSSION

We have developed a highly efficient, low cost and rapid method, CRI-SPA, for parallelized transfer of a genetic feature from a haploid CD strain to any haploid yeast library strain, which is *ura3*. In this report, we have focused on gene transfer into the YKO library, but we note that other libraries like the GFP library (57) and SWAT libraries (15,16) could also be used in CRI-SPA experiments. At the individual level, CRI-SPA mediated gene transfer is highly efficient and can, unlike SGA, be performed without selecting for the genetic feature of interest. At the global level, the efficiency of transfer is also very high and only 62 genes (1.3%) of the library consistently failed to produce any colonies through three screen repeats. As expected, these genes are associated with mating, cell division, and DNA repair, which represent functions that are essential for CRI-SPA (Supplementary Table S7). To this end, we stress that CRI-SPA mediated gene transfer is based on interchromosomal gene conversion. Hence, the donor sequence persists in the diploid state where gene transfer takes place allowing for a long tunable window of opportunity for transfer. This contrasts the scenario in classical gene targeting where the genetic feature is introduced via a linear DNA donor fragment or a self-cleaving plasmid system delivered by a SPA based method (16,58–59). In those cases, the time window where gene transfer can take place is limited to the lifespan of the donor fragment.

We note that CRI-SPA colonies are not clones and that the colonies may contain other cell types than modified recipient cells. Hence CRI-SPA colonies may potentially contain escapers, unmodified recipient-, CD- and diploid cells on the final selection plate. In the betaxanthin experiment when applying selection for the Btx-cassette, these cell types were present in amounts lower than 1% and did not influence the readout significantly. In the corresponding experiment where we did not apply selection for the Btx-cassette, we observed a few cases where large amounts of unmodified recipient strains survived step 2 where diploids are selected. These cases were typically flawed CRI-SPA transfer trials, and we envision that they may form if unusually high amounts of recipient cells are pinned from the starting library plate to the plate where mating takes place. If at the same time, mating is poor, diploid cells may not outcompete unmodified recipient strain on the plate selecting for diploids despite their advantage of selection. We stress that all four trials are rarely tainted by contaminating cells, and if two rounds of CRI-SPA are performed, each with quadruplicate trials, we believe that a recipient library can be reliably covered for ≥99% of the strains. The exceptions will be if the genetic feature of interest introduces a strong fitness disadvantage as this will likely increase the number of undesired other cell types in the CRI-SPA colonies. In these cases, the selection scheme may need further optimization. In the pilot *ade2*Δ*::hphNT1* transfer experiments, where introduction of this mutation comes with a fitness penalty, we added additional adenine to the medium to reduce this effect. Nevertheless, a few colonies were still contaminated by white cells to artificially impact the colony color phenotype. Some of these cells turned out to be diploid recipient YKO cells where only one of the homologous chromosomes had received the *ade2*Δ*::hphNT1* cassette. Although it may be difficult to avoid occasional formation of diploid recipient cells due to endoreduplication, the effect of these cells on the colony phenotype will be minimized if both chromosomes receive the genetic feature of interest. This

can be achieved by optimizing the gRNA to ensure that all target chromosomes are cut in all cells. Consequently, we recommend verifying Cas9-sgRNA cutting efficiency with TAPE, and we have previously observed fold-differences in transformant numbers in the presence and absence of a repair template that exceed 400 (46). Another way to ensure that both sister chromatids of the recipient strain will receive the modification is to prolong the window of opportunity for Cas9 cleavage. This can easily be achieved by extending the duration of the raffinose step where little cell growth takes place. Another source of contaminating cells could in principle be aneuploid cells containing CD chromosomes where the *Kl_URA3* markers were mutated. The forward mutation rate of endogenous *URA3* selecting for 5-FOA resistance is around 5×10^{-8} (60). Assuming that the forward mutation rate of *Kl_URA3* markers are similar to the rate of *URA3,* we find the presence of such cells highly unlikely. Lastly, we note that the betaxanthin phenotype of the hit strains may be caused by other accidental genetic alterations in the recipient strain, which potentially could have been introduced during construction of the gene deletion or during the CRI-SPA procedure. In our Btx-Cassette transfer experiment, we successfully demonstrated that Btx-hit strains maintained their phenotype if they were recreated *de novo* by conventional gene targeting, showing that the phenotype of the hit strain was caused by the specific mutation of the YKO library strain (Figure 6).

In principle, CRI-SPA can be used as a tool to forward the understanding of any biological process in yeast as it allows a genetic feature to be combined with large libraries of other mutations. In this report, we demonstrate in a pilot experiment that CRI-SPA can be used to identify epistatic interactions in a genetic network. Specifically, we identified the epistatic interaction between ade2 and ade3 in a set of other mutant strains (Figure 2). More thoroughly, we demonstrated how CRI-SPA can be used to develop a metabolic-engineering strategy for production of the plant metabolite betaxanthin by identifying target genes that can be manipulated to increase betaxanthin production. This was achieved by using CRI-SPA to transfer a five-genes cassette for betaxanthin production into the YKO library. Our successful transfer of the Btx-cassette to close to all of the strains of the YKO library strains allowed us to provide a systems overview of functions that influence betaxanthin production in yeast. Importantly, our CRI-SPA screen identified genes that are strongly enriched for GO-terms forming functional networks supporting a mechanistic consensus across hits. For example, our analyses showed that deletion of genes, *UBC7* and *SSM4*/*DOA10*, involved in ubiquitin mediated endoplasmic-reticulum-associated protein degradation, ERAD (61), benefit betaxanthin production. In this context it is important to note that the betaxanthin pathway includes a p450 enzyme, CYP76AD1, which is N-terminally linked to the ER membrane. P450s are notoriously difficult to produce in fungal production hosts and it is tempting to speculate that folding of CYP76AD1 in yeast is inefficient and therefore prone to ERAD. In this model, increased betaxanthin levels are the result of the higher CYP76AD1 activity that will materialize in the absence of functional ERAD. From a gene-network perspective, we note that Ubc7 is a ubiquitin conjugating enzyme, which is C-terminally tail-anchored to the ER membrane (62). This fact may explain why *get4*Δ is also scored as a positive hit as Get4 is part of the GET complex that integrates C-terminal-tail-anchored proteins into the ER membrane (63). In another part of the gene network connected by CYP76AD1 we find genes encoding mitochondrial proteins. For example, deletion of *HEM25* negatively impacts betaxanthin production. Hem25 is a mitochondrial transporter facilitating the import of glycine, a precursor for the biosynthesis of heme. Since heme is an essential cofactor of CYP76AD1 (14,64); it is easy to envision that reduced heme availability might impair the activity of CYP76AD1, and in turn betaxanthin production. From another point of view, we also stress that our CRI-SPA screen, in addition to well-characterized genes, was also able to pick up Btx-hit genes with no assigned function. Such gene hits would be difficult, if not impossible, to identify via traditional metabolic-model based hit-finding strategies.

Since our betaxanthin CRI-SPA screen was designed to facilitate future betaxanthin cell factory development, it was important to validate that our CRI-SPA betaxanthin hits identified on solid medium would maintain their phenotype if they were propagated in liquid medium, which is most often used in production. In favor of using CRI-SPA as a valuable starting point in the development of a metabolic engineering strategy, most of the recreated betaxanthin hits influenced betaxanthin production in a manner similar to the original effects observed for the corresponding CRI-SPA colonies. Altogether, CRI-SPA was able to unravel novel physiology impacting betaxanthin production and to identify gene targets that can be used to develop a metabolic engineering strategy for production of betaxanthin, and potentially of L-DOPA derived compounds. For the present, and for other metabolic engineering projects, we envision that by examining gene-interaction networks identified by a CRI-SPA screen, it will be possible to identify mutations that can be advantageously combined. In particular, mutations that are not part of the same functional network may be combined to provide additive, or even synergistic, effects on production of the desired product. We note that not only the positive hits can be used in a future strategy, negative hits may point to limiting functions, which may be improved by overexpressing the hit gene.

Our successful marker-free gene-transfer experiment opens promising avenues for CRI-SPA. Bypassing the need for marker selection sets the stage for the transfer of point mutations as well as the multiplex transfer of several genetic features located at different positions in the genome at once. In another application, CRI-SPA could be used to transfer a genetic trait to libraries with different genetic backgrounds or even closely related species, as long as they can mate with the CD. This is possible as CRI-SPA is independent of meiotic recombination and sporulation. Currently, a limitation of our method is that a recipient library needs to be *ura3* to be compatible with the CRI-SPA selection procedure. This could be bypassed by replacing the *Kl_URA3* markers in the CD strain by another counter-selectable marker like *amdS* (65) to produce an even more flexible version of CRI-SPA.

To conclude, we have demonstrated that CRI-SPA is a powerful high-throughput data generation tool opening up

a range of new exciting applications both in fundamental genetics and in genome engineering.

DATA AVAILABILITY

The top hits, bottom hits and failing genes lists are available as Supplementary Tables 5, 6 and 7, respectively. A sample of 13 screen images, the raw datasets obtained from such images for each screen, the individual and the corrected + filtered datasets are all available on our CRI-SPA github repository (https://github.com/pc2912/CRI-SPA_repo). All scripts used in the analysis and for the drawing of the figures of this work are also available on the CRI-SPA github repository (permanent doi: https://doi.org/10.5281/zenodo.8160760).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Rodney Rothstein, Columbia University (NY, USA) for the strains W8164-2B and W8164-2C; Irina Borodina, DTU (Denmark), for the plasmids pCfB2311 and pCfB2312; and Christina S. Nødvig for plasmid pXI-2-ARO4*-ARO7*. We also thank the members of the Mortensen lab for helpful discussions.
Author contributions: P.C., H.O., H.C., M.L. and U.H.M., designed research. P.C., H.O., H.C., M.L.J, M.vcB., B.J.S. and T.S. performed research. P.C., H.O., J.-M.D. and U.H.M. analyzed data. N.S., M.K.J. and U.H.M. supervised the research. P.C., H.O. and U.H.M wrote the paper.

FUNDING

European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie [722287 (PAcMEN) to U.H.M., M.K.J., J.M.D.]; Danish National Advanced Technology Foundation [087-2012-3 to U.H.M.]; Villum Foundation and the Danish National Research Foundation [DNRF115 to M.L.]; Novo Nordisk Foundation Bioscience Ph.D. Programme [NNF19SA0035438 to P.C.]; Fermentation-based Biomanufacturing Initiative [NNF17SA0031362 to N.S.]. Funding for open access charge: The Danish National Advanced Technology Foundation [087-2012-3].
Conflict of interest statement. None declared.

REFERENCES

1. Goffeau,A., Barrell,B.G., Bussey,H., Davis,R.W., Dujon,B., Feldmann,H., Galibert,F., Hoheisel,J.D., Jacq,C., Johnston,M. *et al.* (1996) Life with 6000 genes. *Science*, **274**, 546–567.
2. Cherry,J.M., Ball,C., Weng,S., Juvik,G., Schmidt,R., Adler,C., Dunn,B., Dwight,S., Riles,L., Mortimer,R.K. *et al.* (1997) Genetic and physical maps of *saccharomycescerevisiae*. *Nature*, **387**, 67.
3. DiCarlo,J.E., Norville,J.E., Mali,P., Rios,X., Aach,J. and Church,G.M. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.*, **41**, 4336–4343.
4. Replogle,J.M., Norman,T.M., Xu,A., Hussmann,J.A., Chen,J., Cogan,J.Z., Meer,E.J., Terry,J.M., Riordan,D.P., Srinivas,N. *et al.* (2020) Combinatorial single-cell crispr screens by direct guide rna capture and targeted sequencing. *Nat. Biotechnol.*, **38**, 954–961.
5. Lian,J., Schultz,C., Cao,M., HamediRad,M. and Zhao,H. (2019) Multi-functional genome-wide crispr system for high throughput genotype–phenotype mapping. *Nat. Commun.*, **10**, 5794.
6. Sharon,E., Chen,S.-A.A., Khosla,N.M., Smith,J.D., Pritchard,J.K. and Fraser,H.B. (2018) Functional genetic variants revealed by massively parallel precise genome editing. *Cell*, **175**, 544–557.
7. Bao,Z., HamediRad,M., Xue,P., Xiao,H., Tasan,I., Chao,R., Liang,J. and Zhao,H. (2018) Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision. *Nat. Biotechnol.*, **36**, 505–508.
8. Han,K., Jeng,E.E., Hess,G.T., Morgens,D.W., Li,A. and Bassik,M.C. (2017) Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.*, **35**, 463–474.
9. Wong,A.S., Choi,G.C., Cui,C.H., Pregernig,G., Milani,P., Adam,M., Perli,S.D., Kazer,S.W., Gaillard,A., Hermann,M. *et al.* (2016) Multiplexed barcoded CRISPR-Cas9 screening enabled by combigem. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 2544–2549.
10. Garst,A.D., Bassalo,M.C., Pines,G., Lynch,S.A., Halweg-Edwards,A.L., Liu,R., Liang,L., Wang,Z., Zeitoun,R., Alexander,W.G. *et al.* (2017) Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.*, **35**, 48–55.
11. Sadhu,M.J., Bloom,J.S., Day,L., Siegel,J.J., Kosuri,S. and Kruglyak,L. (2018) Highly parallel genome variant engineering with CRISPR–CAS9. *Nat. Genet.*, **50**, 510–514.
12. Bock,C., Datlinger,P., Chardon,F., Coelho,M.A., Dong,M.B., Lawson,K.A., Lu,T., Maroc,L., Norman,T.M., Song,B. *et al.* (2022) High-content crispr screening. *Nat. Rev. Methods Primers*, **2**, 8.
13. Shalem,-O., Sanjana,N.E. and Zhang,F. (2015) High-throughput functional genomics using crispr–cas9. *Nat. Rev. Genet.*, **16**, 299–311.
14. Savitskaya,J., Protzko,R.J., Li,F.-Z., Arkin,A.P. and Dueber,J.E. (2019) Iterative screening methodology enables isolation of strains with improved properties for a facs-based screen and increased l-dopa production. *Sci. Rep.*, **9**, 5815.
15. Weill,U., Yofe,I., Sass,E., Stynen,B., Davidi,D., Natarajan,J., Ben-Menachem,R., Avihou,Z., Goldman,O., Harpaz,N. *et al.* (2018) Genome-wide swap-tag yeast libraries for proteome exploration. *Nat. Methods*, **15**, 617–622.
16. Yofe,I., Weill,U., Meurer,M., Chuartzman,S., Zalckvar,E., Goldman,O., Ben-Dor,S., Schutze,C., Wiedemann,N., Knop,M. *et al.* (2016) One library to make them all: streamlining the creation of yeast libraries via a swap-tag strategy. *Nat. Methods*, **13**, 371–378.
17. Tong,A.H.Y., Evangelista,M., Parsons,A.B., Xu,H., Bader,G.D., Pagé,N., Robinson,M., Raghibizadeh,S., Hogue,C.W., Bussey,H. *et al.* (2016) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, **294**, 2364–2368.
18. Kuzmin,E., Rahman,M., VanderSluis,B., Costanzo,M., Myers,C.L., Andrews,B.J. and Boone,C. (2021) τ -sga: synthetic genetic array analysis for systematically screening and quantifying trigenic interactions in yeast. *Nat. Protoc.*, **16**, 1219–1250.
19. Costanzo,M., VanderSluis,B., Koch,E.N., Baryshnikova,A., Pons,C., Tan,G., Wang,W., Usaj,M., Hanchard,J., Lee,S.D. *et al.* (2016) A global genetic interaction network maps a wiring diagram of cellular function. *Science*, **353**, aaf1420.
20. Lichten,M. (2001) Meiotic recombination: breaking the genome to save it. *Curr. Biol.*, **11**, R253–R256.
21. Adames,N.R., Gallegos,J.E. and Peccoud,J. (2019) Yeast genetic interaction screens in the age of CRISPR/Cas. *Curr. Genet.*, **65**, 307–327.
22. Jaffe,M., Sherlock,G. and Levy,S.F. (2017) iseq: a new double-barcode method for detecting dynamic genetic interactions in yeast. *G3: Genes Genomes Genetics*, **7**, 143–153.
23. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-rna–guided dna endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
24. Reid,-R.J., Gonźalez-Barrera,S., Sunjevaric,I., Alvaro,D., Ciccone,S., Wagner,M. and Rothstein,R. (2011) Selective ploidy ablation, a high-throughput plasmid transfer protocol, identifies new genes affecting topoisomerase i–induced dna damage. *Genome Res.*, **21**, 477–486.
25. DeLoache,W.C., Russ,Z.N., Narcross,L., Gonzales,A.M., Martin,V.J. and Dueber,J.E. (2015) An enzyme-coupled biosensor enables

(s)-reticuline production in yeast from glucose. *Nat. Chem. Biol.*, **11**, 465–471.
26. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., eronneau,S.V., Dow,S., Lucau-Danila,A., Anderson,K., Andŕe,B. *et al.* (2002) Functional profiling of the *Saccharomycescerevisiae* genome. *Nature*, **418**, 387–391.
27. Reid,R.J., Sunjevaric,I., Voth,W.P., Ciccone,S., Du,W., Olsen,A.E., Stillman,D.J. and Rothstein,R. (2008) Chromosome-scale genetic mapping using a set of 16 conditionally stable *Saccharomycescerevisiae* chromosomes. *Genetics*, **180**, 1799–1808.
28. Sherman,F., Fink,G. and Hicks,J. (1986) Laboratory course manual for methods in yeast genetics. cold spring harbor, ny: cold spring harbor lab. Press. ’
29. Geu-Flores,F., Nour-Eldin,H.H., Nielsen,M.T. and Halkier,B.A. (2007) User fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple pcr products. *Nucleic Acids Res.*, **35**, e55.
30. Nour-Eldin,H.H., Hansen,B.G., Nørholm,M.H., Jensen,J.K. and Halkier,B.A. (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning pcr fragments. *Nucleic Acids Res.*, **34**, e122.
31. Gietz,R.D. and Schiestl,R.H. (2007) High-efficiency yeast transformation using the liac/ss carrier DNA/PEG method. *Nat. Protoc.*, **2**, 31–34.
32. Jensen,N.B., Strucko,T., Kildegaard,K.R., David,F., Maury,J., Mortensen,U.H., Forster,J., Nielsen,J. and Borodina,I. (2014) Easyclone: method for iterative chromosomal integration of multiple genes *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **14**, 238–248.
33. Stovicek,V., Borodina,I. and Forster,J. (2015) CRISPR–cas system enables fast and simple genome editing of industrial *Saccharomycescerevisiae* strains. *Metab. Eng. Commun.*, **2**, 13–22.
34. Rothstein,R.J. (1983) One-step gene disruption in yeast. *Methods Inenzymol.*, **101**, 202–211.
35. Kamrad,S., Rodriguez-Lopez,M., Cotobal,C., Correia-Melo,C., Ralser,M. and Bähler,J. (2020) Pyphe, a python toolbox for assessing microbial growth and cell viability in high-throughput colony screens. *eLife*, **9**, e55160.
36. Baryshnikova,-., Costanzo,M., Kim,Y., Ding,H., Koh,J., Toufighi,K., Youn,J.-Y., Ou,J., San Luis,B.-J., Bandyopadhyay,S. *et al.* (2010) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods*, **7**, 1017–1024.
37. Collins,S.R., Schuldiner,M., Krogan,N.J. and Weissman,J.S. (2006) A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.*, **7**, R63.
38. Klopfenstein,D., Zhang,L., Pedersen,B.S., Ramírez,F., Warwick Vesztrocy,A., Naldi,A., Mungall,C.J., Yunes,J.M., Botvinnik,O., Weigel,M. *et al.* (2018) Goatools: a python library for gene ontology analyses. *Sci. Rep.*, **8**, 10872.
39. Gene Ontology Consortium (2006) The gene ontology (go) project in 2006. *Nucleic Acids Res.*, **34**, D322–D326.
40. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
41. Mikkelsen,M.D., Buron,L.D., Salomonsen,B., Olsen,C.E., Hansen,B.G., Mortensen,U.H. and Halkier,B.A. (2012) Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.*, **14**, 104–111.
42. Baker Brachmann,C., Davies,A., Cost,G.J., Caputo,E., Li,J., Hieter,P. and Boeke,J.D. (2098) Designer deletion strains derived from *Saccharomycescerevisiae* s288c: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, **14**, 115–132.
43. Filippo,J.S., Sung,P. and Klein,H. (2008) Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.*, **77**, 229–257.
44. Jinks-Robertson,-S. and Petes,T.D. (2021) Mitotic recombination in yeast: what we know and what we don’t know. *Curr. Opin. Genet. Dev.*, **71**, 78–85.
45. Johnston,M., Flick,J.S. and Pexton,T. (1994) Multiple mechanisms provide rapid and stringent glucose repression of gal gene expression in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **14**, 3834–3841.
46. Vanegas,K.G., Lehka,B.J. and Mortensen,U.H. (2017) Switch: a dynamic crispr tool for genome engineering and metabolic pathway control for cell factory construction in *Saccharomyces cerevisiae*. *Microb. Cell Fact.*, **16**, 25.
47. Wintersberger,U., Kühne,C. and Karwan,A. (1995) Scp160p, a new yeast protein associated with the nuclear membrane and the endoplasmic reticulum, is necessary for maintenance of exact ploidy. *Yeast*, **11**, 929–944.
48. Smirnov,M., Smirnov,V., Budowsky,E., Inge-Vechtomov,S. and Serebrjakov,N. (1967) Red pigment of adenine-deficient yeast *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.*, **27**, 299–304.
49. Roman,H. (1956) Red color in ade2 but with epistasis to ade3. *Cold Spring Harbor Symp. Quant. Biol.*, **21**, 175.
50. Chaudhuri,B., Ingavale,S. and A.,K. (1997) Bachhawat. Apd1+, a gene required for red pigment formation in *ADE6* mutants of *Schizosaccharomyces pombe,* encodes an enzyme required for glutathione biosynthesis: a role for glutathione and a glutathione-conjugate pump. *Genetics*, **145**, 75–83.
51. Hughes,T.R., Roberts,C.J., Dai,H., Jones,A.R., Meyer,M.R., Slade,D., Burchard,J., Dow,S., Ward,T.R., Kidd,M.J. *et al.* , (2000) Widespread aneuploidy revealed by dna microarray expression profiling. *Nat. Genet.*, **25**, 333–337.
52. Gelin-Licht,R., Paliwal,S., Conlon,P., Levchenko,A. and Gerst,J.E. (2012) Scp160-dependent mRNA trafficking mediates pheromone gradient sensing and chemotropism in yeast. *Cell Rep.*, **1**, 483–494.
53. Guo,M., Aston,C., Burchett,S.A., Dyke,C., Fields,S., Rajarao,S.J.R., Uetz,P., Wang,Y., Young,K. and Dohlman,H.G. (2003) The yeast G protein α subunit gpa1 transmits a signal through an rna binding effector protein SCP160. *Mol. Cell*, **12**, 517–524.
54. Luttik,M., Vuralhan,Z., Suir,E., Braus,G., Pronk,J. and Daran,J. (2008) Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: quantification of metabolic impact. *Metab. Eng.*, **10**, 141–153.
55. Gandía-Herrero,F., Garcıa-Carmona,F. and Escribano,J., (2005) A novel method using high-performance liquid chromatography with fluorescence detection for the determination of betaxanthins. *J. Chromatogr. A*, **1078**, 83–89.
56. Jessop-Fabre,M.M., Jakoˇcīunas,T., Stovicek,V., Dai,Z., Jensen,M.K., Keasling,J.D. and Borodina,I. (2016) Easyclone-markerfree: a vector toolkit for markerless integration of genes into *Saccharomyces cerevisiae* via CRISPR-Cas9. *Biotechnol. J.*, **11**, 1110–1117.
57. Huh,W.-K., Falvo,J.V., Gerke,L.C., Carroll,A.S., Howson,R.W., Weissman,J.S. and O’Shea,E.K. (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686–691.
58. Berry,L.K., Thomas,G.H. and Thorpe,P.H. (2020) Cats: cas9-assisted tag switching. A high-throughput method for exchanging genomic peptide tags in yeast. *BMC Genomics [Electronic Resource]*, **21**, 221.
59. Roy,K.R., Smith,J.D., Vonesch,S.C., Lin,G., Tu,C.S., Lederer,A.R., Chu,A., Suresh,S., Nguyen,M., Horecka,J. *et al.* (2018) Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat. Biotechnol.*, **36**, 512–520.
60. Lang,G.I. and Murray,A.W. (2008) Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics*, **178**, 67–82.
61. Heard,E. (2010) Editorial— Seminars in cell & developmental biology. In: *Seminars in Cell & Developmental Biology*. Academic Press, Vol. **21**, p. 185.
62. Sommer,T. and Jentsch,S. (1993) A protein translocation defect linked to ubiquitin conjugation at the endoplasmic reticulum. *Nature*, **365**, 176–179.
63. Farkas,A. and Bohnsack,K.E. (2021) Capture and delivery of tail-anchored proteins to the endoplasmic reticulum. *J. Cell Biol.*, **220**, e202105004.
64. Hatlestad,G.J., Sunnadeniya,R.M., Akhavan,N.A., Gonzalez,A., L.Goldman,I., McGrath,J.M. and Lloyd,A.M., (2012) The beet r locus encodes a new cytochrome p450 required for red betalain production. *Nat. Genet.*, **44**, 816–820.
65. Solis-Escalante,D., Kuijpers,N.G., Nadine,B., Bolat,I., Bosman,L., Pronk,J.T., Daran,J.M. and Pascale,D.-L. (2013) AmdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **13**, 126–139.

Chapter 4

Manuscript 2: Microbial cell factory optimisation using genome-wide host-pathway interaction screens

Foreword

The following manuscript has been deposited on BiorXiv: https://doi.org/10.1101/2023.08.30.555557

Microbial cell factory optimisation using genome-wide host-pathway interaction screens

Paul Cachera[1], Nikolaj Can Kurt[2], Andreas Røpke[2], Tomas Strucko[2], Uffe H. Mortensen[2], Michael K. Jensen[1,#]

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

[2] Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, Denmark

[#] = to whom correspondence should be addressed: Michael K. Jensen mije@biosustain.dtu.dk

Abstract

The ubiquity of genetic interactions in living cells challenges the concept of parts orthogonality, which is a cornerstone of synthetic biology. Parts, such as heterologously expressed genes, draw from shared pools of limited cellular resources and interactions between parts themselves and their host are inevitable. Instead of trying to eliminate or disregard these interactions, we propose to leverage them to promote desirable phenotypes. We recently described CRI-SPA, a method for high-throughput genome-wide gene delivery and screening of host:pathway interactions in *Saccharomyces cerevisiae*. In this study, we combine this method with biosensor-based high-throughput screening and high-density colony image analysis to identify lead engineering targets for optimising *cis*-*cis*-muconic acid (CCM) production in yeast cell factories. Using the biosensor screen, we phenotype >9,400 genotypes for their interaction with the heterologously expressed CCM biosynthesis pathway, including both gene knock-out and overexpression, and identify novel metabolic targets belonging to sulphur assimilation and methionine synthesis, as well as cellular redox homeostasis, positively impacting CCM biosynthesis by up to 280%. Our genome-wide exploration of host pathway interaction opens novel strategies for the metabolic engineering of yeast cell factories.

Introduction

The field of synthetic biology aims at engineering biological functions *de novo*, either *in vitro* or *in vivo*, by cataloguing individual biological components as orthogonal parts (Canton et al., 2008). The notion that cellular systems can be packaged into independent modules is a useful abstraction, which facilitates studying biological complexity (Hartwell et al., 1999). This reductionist framework is enticing because it is conceptually simple and successfully rooted at the core of many engineering disciplines (Lazebnik, 2002). Bottom-up designs in living cells have been undeniably successful in the building of complex transcriptional logics (Friedland et al., 2009; Nielsen et al., 2016; Zong et al., 2017), and orthogonal transcription and translation of non-canonical amino acids (d'Aquino et al., 2018). However, such successes are built upon a thorough understanding of the underlying molecular processes and this understanding is lacking for numerous cellular mechanisms even in model organisms (Wood et al., 2019)

Reducing the complexity of interactions between heterologous gene products and their host to a handful of inputs-outputs is likely an oversimplification. Indeed, heterologously produced enzymes might interfere with the host cellular processes by e.g. exhausting pools of cofactors and energy, catalysing side reactions and producing toxic intermediates (Boo et al., 2019; Dahl et al., 2013; Sun et al., 2017; Wu et al., 2016). This challenge escalates dramatically when the number of coding sequences and edits that current metabolic engineering efforts currently introduce into cell factories is considered (Zhang et al., 2022).

The concept of host-aware synthetic biology has gained momentum in the past 10 years with a number of efforts addressing interactions between the synthetic system and the host (Boo et al., 2019; Cardinale and Arkin, 2012; de Lorenzo, 2011). For example, studies have quantified and attempted to reduce the heterologous burden of synthetic systems on the host (Borkowski et al., 2018; Ceroni et al., 2015; Gyorgy et al., 2015; Shopera et al., 2017; Weiße et al., 2015). An alternative approach is to reduce the cross talk between parts as well as between parts and the host, a strategy known as 'isolation' (Darlington et al., 2018; Lou et al., 2012; Meyer et al., 2015; Segall-Shapiro et al., 2014). A burgeoning branch of synthetic biology goes as far as proposing to fully isolate human-made mechanisms through a complete orthogonal dogma (Costello and Badran, 2021; Liu et al., 2018).

Instead of eliminating genetic interactions, we reason that they may serve an engineering task. Indeed, the systematic combination of double KO in yeast revealed that, in rare instances (≈1.5%), the interaction of two gene KOs could yield an unexpected increase in fitness (Costanzo et al., 2016, 2010). This hints that sporadic phenotypic maximas hidden in the interaction landscape are awaiting to be identified and exploited. This idea is supported by adaptive laboratory evolution and multiplexed CRISPR-based screens which find that combinations of edits often act synergistically to yield a sought-for phenotype which cannot be reduced to the sum of its components (Bao et al., 2018; Lian et al., 2017; Wang et al., 2020).

The identification of rare phenotype maxima in the interaction landscape requires the means to generate and analyse strain variants at high throughput. We recently described CRI-SPA, a method capable of delivering a genetic feature to an arrayed yeast library within a week (Cachera et al., 2023). In its first version, we used CRI-SPA to screen the yeast knowck-out (YKO) library for gene knockouts interacting with the biosynthesis of the plant pigment betaxanthin. The workflow made use of the yellow phenotype produced by betaxanthin to extract a productivity readout with image analysis. Image analysis is arguably the simplest and most inexpensive high-throughput readout strategy. Unfortunately, only a minority of value-added products, such as betaxanthin and carotenoids, directly produce a visible phenotype (Zeng et al., 2020). Biosensors, which translate a chemical input into a measurable output, typically fluorescence, widen the range of products and cellular processes which can be measured and quantified (d'Oelsnitz et al., 2022; Rogers et al., 2016). A biosensor can be expressed in the producing strain and scale to any throughput with no additional cost. In addition, a wide collection of biosensors have been developed over the past decades, opening their application to a large variety of products (Koch et al., 2019).

Here, we couple CRI-SPA to the biosensing of the otherwise inconspicuous bio-plastic platform chemical *cis-cis*-muconic acid (CCM) with the aim of identifying novel

host:pathway interactions in *S. cerevisiae.* Since CCM was first synthesised in yeast a decade ago (Curran et al., 2013b), numerous metabolic engineering efforts have attempted to increase its production by improving the availability of shikimate pathway precursors or by relieving the enzymatic bottleneck of the rate-limiting conversion of protocatechuic acid (PCA) to catechol (**Fig. 1a**), (Brückner et al., 2018; Curran et al., 2013a; Suástegui et al., 2017; Weber et al., 2012, 2017). CCM is the ideal test-bed for our study because identification of rational engineering strategies to increase its titers likely have reached a plateau. On the other hand, the engineering of biosensors for CCM (Skjoedt et al., 2016; Snoek et al., 2020) and for shikimate pathway end products (Leavitt et al., 2016) have supported a number of screens and directed evolution studies (Jensen et al., 2021; Leavitt et al., 2017; Snoek et al., 2018; Wang et al., 2021, 2020). Biosensor-coupled directed evolution can efficiently isolate strains with higher production capacity. However, disentangling the mechanisms behind these improvements and their reverse engineering can be complicated by the number of mutations which need to be tested individually or in combination (Mundhada et al., 2017; Wang et al., 2020).

To open new engineering avenues for CCM synthesis, we sought to screen for beneficial host:pathway interactions resulting in an increased CCM accumulation in *S. cerevisiae.* We used CRI-SPA to deliver a cassette encoding for the synthesis and reporting of CCM to the strains of the genome-wide OverExpression (OEx) and Knock-Out (KO) libraries. We measured CCM biosensing directly from colonies growing on agar plates enabling high-density array image analysis and identified known and novel cellular functions interacting with CCM synthesis. Importantly, many of the positive hits identified in the screen were subsequently validated and shown to outperform the reference strain in liquid medium.

Results

Biosensor-assisted CCM quantification based on image analysis

To enable a genetically-encoded biosensor to report a proxy for CCM formation, we engineered the CRI-SPA donor strain (Cachera et al., 2023) with an expression cassette encoding the CCM biosynthesis pathway and the CCM biosensor (**Fig. 1a; Supplementary Figure S1**). Engineering of the CRI-SPA donor strain was initiated by building a single 19kb DNA cassette able to both synthesise and report CCM (**Fig. 1a**). Subsequently, the CCM cassette is then designed to transfer from the donor strain to yeast libraries using CRI-SPA, (Giaever et al., 2002; Weill et al., 2018; Yofe et al., 2016), and edited strains of the library screened for fluorescent signals as a proxy for CCM synthesis (**Fig. 1b**).

Specifically, for CCM synthesis, we refactored the previously described biosynthetic pathway (Skjoedt et al., 2016; Snoek et al., 2020) by replacing all repetitive promoters and

terminators with characterised strong equivalents (Curran et al., 2013a; Zhang et al., 2020) (**Suppl. Figure S1**). This minimises the risk of repeat-recombination events within the pathway that may result in gene rearrangement or gene copy loss during regular strain propagation or during CRI-SPA-mediated cassette transfer. We also engineered an additional donor strain expressing the decarboxylase AroY-B variant (AroY-B_P146T), which we previously evolved for optimized conversion of PCA to catechol (Jensen et al., 2021**, Fig. 2a**). For CCM detection, an improved version of the BenM biosensor (BenM17_D08, henceforth referred to as BenM) was encoded in the 19 kb cassette (Snoek et al., 2020).

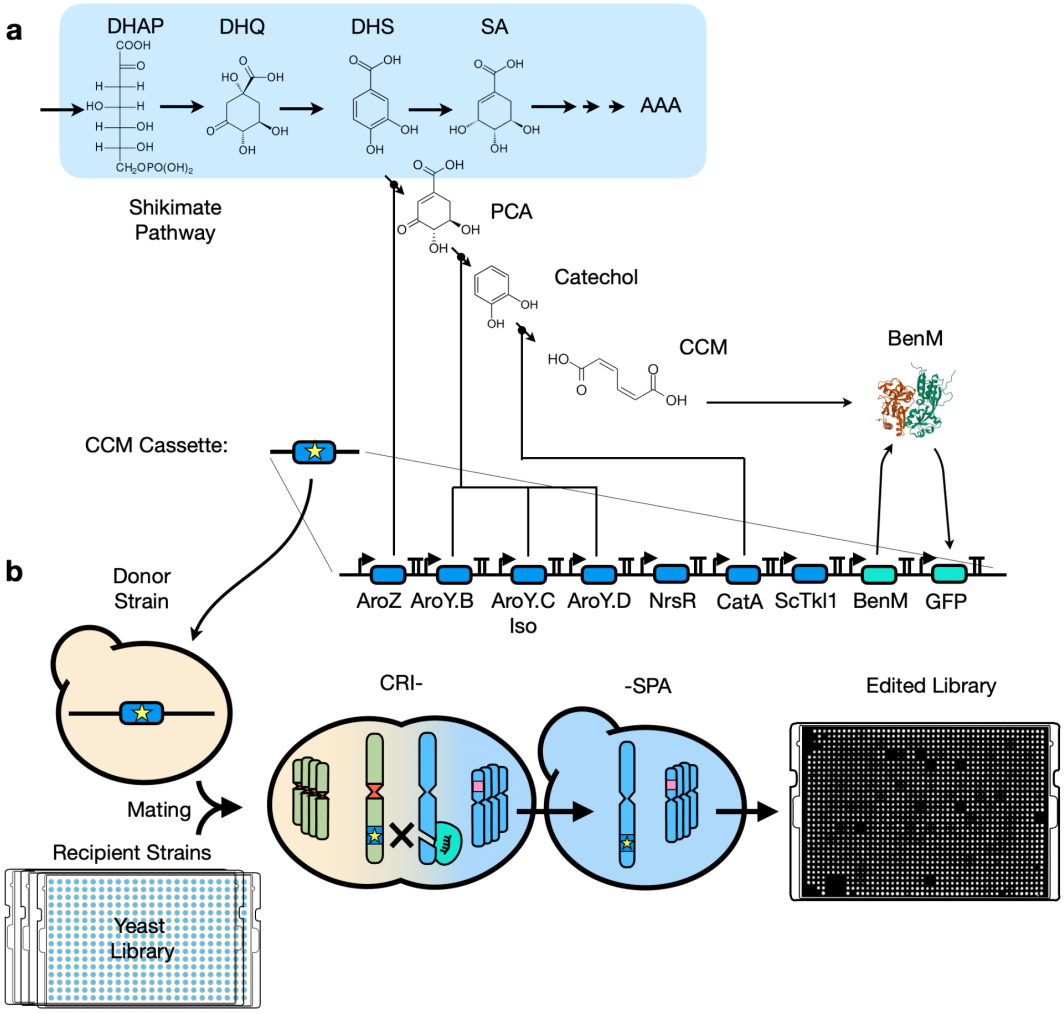


Fig. 1. Overview of the design of the CRI-SPA screen for identification of host genes interacting with CCM synthesis. **a**) Schematic of the synthesis and reporting of CCM by the CCM cassette. Abbreviations: AroZ: *Klebsiella pneumoniae* DHS dehydratase. AroY.B-AroY.Ciso-AroyD: *K. pneumoniae* PCA decarboxylase cluster (iso=isoform). NrsR: Nourseothricin Resistance, CatA: *Acinetobacter baylyi* catechol 1,2-dioxygenase DAHP: 3-Deoxy-D-arabinoheptulosonate 7-phosphate, ScTKL1: *S. cerevisiae* Transketolase, BenM:

engineered CCM biosensor. DHQ: 3-Dehydroquinic acid, DHS: 3-dehydroshikimate, SA: shikimic acid, AAA: Aromatic Amino Acids. Refer to **Suppl. Fig. S1** for all parts in the cassette and **Suppl. Table S1** for the sourcing of these parts. **b**) The CCM cassette is inserted in the CRI-SPA donor strain and delivered to all strains of a yeast library by CRI-SPA. Final colony fluorescence can be extracted with image analysis as a proxy for quantifying CCM titers.

For CRI-SPA-mediated editing, we inserted the CCM cassettes in the XII-5 donor site of the CRI-SPA donor strain of both mating types (Cachera et al., 2023). For the *Mat-**a*** donor strains, this resulted in strain CRI-SPA Donor strain CD1 (*Mat-**a*** with expression of WT *AroY-B*) and CD2 (*Mat-**a*** with optimised AroY-B_P146T). For the *Mat-**alpha*** donor strains, the cassettes were chromosomally inserted and the expression cassette encoding nourseothricin resistance was switched to kanamycin resistance, resulting in CD3 (*Mat-alpha* with WT AroY-B) and CD4 (*Mat-alpha* with optimised AroY-B_P146T). In parallel to these CDs, we built two control strains, BY1 and BY2 harbouring the two pathway variants with WT AroY-B or the optimized AroY-B_P146T in background BY4741. BY4741 is commonly used in the making of yeast libraries (Giaever and Nislow, 2014; Weill et al., 2018; Yofe et al., 2016)(**Fig. 2a**). BY1 and BY2 thus serve as controls identical to the final edited library that the CRI-SPA screens will produce. Finally, to test the inducibility of BenM in the library background, we also built BY0 a strain harbouring the CCM biosensor but without the CCM biosynthetic pathway.

Next, we validated the production of CCM and its biosensing in the donor strain and the library controls, and found that the refactored pathways produced CCM at levels in agreement with previous work (Jensen et al., 2021)(**Fig. 2b**). We also detected the expected titer increase caused by the optimised AroY.B decarboxylase (WT: 138.7 mg/L, optimised: 168.0 mg/L, two-sided t-test, P-value = 0.012). Furthermore, fluorescence analysis measuring production of the GFP reporter confirmed the inducibility of the biosensor by CCM, albeit with a lower dynamic output range (2.5x) than previously reported (Snoek et al., 2020)(**Fig. 2b**). Encouragingly, the biosensor was able to detect the modest, yet insignificant, increase in CCM production caused by the optimised AroY.B compared to the WT decarboxylase, indicating that the level of CCM produced by the CCM biosynthetic pathway fell within the operational range of detection of the biosensor (**Fig. 2b**).

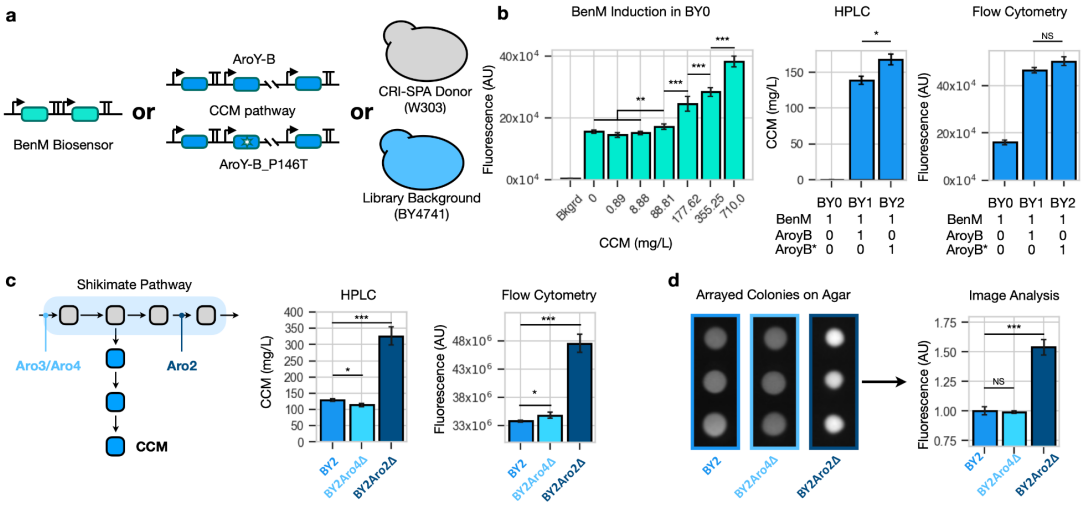


Fig. 2. Characterization of CCM production and biosensing in CRI-SPA engineered yeast. a) The two variants of the CCM cassette, harbouring either the WT AroY-B decarboxylase or the optimised variant AroY-B_P146T, were introduced in the CRI-SPA donor strain and the BY4741 background. **b**) Left: Induction of the CCM biosensor in strain BY0. Centre: CCM titers for strains BY0, BY1 and BY2. Right: fluorescence measured by flow cytometry for strains BY0, BY1 and BY2. **c**) KO of genes *ARO2* and *ARO4* as positive and negative controls, respectively. Left: the reactions catalysed by Aro2 and Aro4 are upstream and downstream of the CCM branch. Centre: CCM synthesis by *aro2*Δ and *aro4*Δ KO strains. Right: Fluorescence of *aro2*Δ and *aro4*Δ strains. **d**) Changes in CCM titers cause a fluorescent change quantifiable directly from agar colonies. Fluorescence is expressed in fold changes relative to BY2. For b) and d), bars show the mean of at least three experimental replicates. For b) and c), replicates are technical. Biological replicates of the experiments can be found **Suppl. Fig. S2**. For all graphs, significance was assessed with a two-sided t-test *<0.05, **<0.01, ***<0.001. AU: arbitrary units.

To further test the ability of the biosensor to detect differences in CCM biosynthesis caused by gene edits in CRI-SPA strains, we built additional control strains that are expected to show increased or decreased CCM synthesis. Here, we deleted *ARO4* encoding DAHP synthase and *ARO2* encoding chorismate synthase, which act up- and downstream of the branchpoint to the heterologous CCM pathway, respectively (Jones et al., 1991; Künzler et al., 1992). The genes were deleted in BY2, the most productive control strain variant, resulting in strains BY2Aro4Δ and BY2Aro2Δ which were cultivated in liquid Synthetic Complete (SC) medium for 72 h (**Fig. 2c**). For the *aro2*Δ strain, we observed more than 2-fold increase (p-value < 0.001) in CCM titers. The *aro4*Δ strain displayed only a modest

reduction in CCM production relative to the wild-type strain (WT: 129.9 mg/L*, aro4Δ:* 113.5 mg/L, p = 0.034). This is probably due to the activity of redundant *ARO3* which can maintain the flux through the shikimate pathway even in the absence of *ARO4* (Curran et al., 2013b). Importantly, quantification of the GFP reporter by flow cytometry of liquid cultures of strains BY2, BY2Aro4Δ and BY2Aro2Δ correlated with the CCM levels obtained by HPLC, confirming that the biosensor was able to report changes in CCM synthesis caused by single gene deletions in CRI-SPA strains (**Fig. 2c**).

To match the high-density agar array format of CRI-SPA, we investigated the possibility of reporting CCM productivity from BY2, BY2Aro4Δ and BY2Aro2Δ directly from colonies growing on agar. In this case, image analysis of colonies of strain *aro2*Δ detected a 56% increase (p-value < 0.001) in colony fluorescence intensity compared to the reference strain (**Fig. 2d**). Importantly, the fluorescence intensity observed for BY2, BY2Aro4Δ and BY2Aro2Δ again correlated with the analytical quantification performed by HPLC (**Fig. 2c**).

In summary, we demonstrated that when our CCM biosensor is implemented in a setup mimicking CRI-SPA output strains, it is able to differentiate host KOs causing an increase in CCM synthesis under screening conditions.

High-density image analysis of >9,400 genotypes

Next, we sought to employ this framework to screen for genome-wide host:pathway interactions for all genes present in the YKO and OEx libraries. For this purpose, we employed CD2 to deliver the CCM cassette to all strains of the YKO and OEx libraries, interrogating fluorescence signals as a proxy for CCM accumulation in a total of >9,400 arrayed mutant strains. As expected, the arrayed colonies edited by CRI-SPA were fluorescent, hinting to a successful transfer of the 19kb CCM cassette into the libraries (**Suppl. Fig. S3a**). We also observed that positional artefacts, which are known to affect colony size in high-density arrays, also affected the colony fluorescence in our screen (Baryshnikova et al., 2010**, Suppl. Fig. S3b-c**). To mitigate this phenomenon, we applied the data processing methods developed by the Boone and Andrews labs to correct for positional artefacts and to remove colony outliers from our data (Wagih et al., 2013).

Following this processing step, we ranked the genes in the two libraries according to their corrected fluorescence. Here, we found that the screen of the KO library had a stronger signal-to-noise ratio than the OEx screen (**Fig. 3a**, **Suppl. Fig. S3a**). The *Z* scores for the first and last 2% quantiles were 9.45 and -15.64 for the KO screen, respectively, but only 4.21 and -4.83 for the OEx screen. This difference in dynamic range was also qualitatively visible on raw screen pictures, where the CRI-SPA colonies on the KO library plates displayed more variations in fluorescence intensity compared to the CRI-SPA colonies on the OEx library plates (**Suppl. Fig. S3a**). Yet, and most importantly, the fluorescence signal was visible on quadruplicate colony

images recovered from the images of both screens (**Fig. 3a-b**). Furthermore, repeating the YKO library screen, we showed that the fluorescent signal obtained from the CRI-SPA colonies were reproducible (PCC = 0.71, p<0.001, **Fig. 3a, insert**). Combining the data produced by the two YKO screen repeats and the single OEx screen (see materials and methods), our final dataset quantified the effects of 4,596 gene KOs (97.8% of the YKO library) and of 4,983 genes overexpressed (99.0% of the OEx library) on the synthesis of CCM.

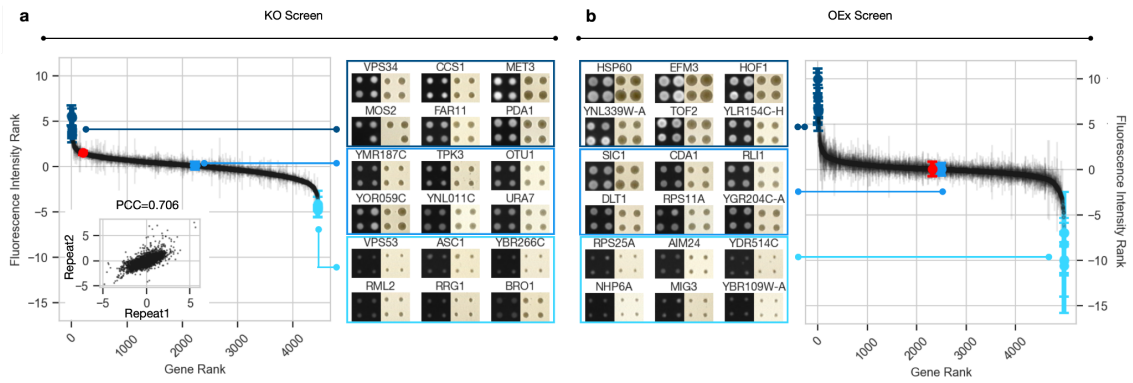


Fig. 3. Ranking of host:pathway interactions for gene KO and overexpression. a) Left: Ranking the fluorescence intensity scores for 4,452 gene KO obtained with the first repeat of the screen on the YKO library. Inset shows reproducibility of the fluorescence intensity scores obtained for each mutant strain in the gene YKO library between two independent trials. Right: Fluorescent and non-fluorescent images of colonies from the same plate sampled at the indicated ranks. **b**) Right: Ranking the fluorescent intensity scores for 4,983 gene OEx strains obtained with the OEx strain library. Left: Fluorescent and non-fluorescent images of colonies sampled at the indicated ranks. Each point is the mean fluorescence of three or more independent CRI-SPA colonies, the error bar is the standard deviation. Red points indicate the WT library strain edited by CRI-SPA.

CCM biosensor identifies known host:pathway interactions

Encouragely, we found that the ranking produced by the KO and OEx fluorescence datasets captured the beneficial effects of knocking-out *ZWF1* and overexpressing of *PAD1*, two common targets for CCM overproduction in *S. cerevisiae* (Curran et al., 2013b; Weber et al., 2012). Indeed, deletion of *ZWF1*, which encodes a cytoplasmic glucose-6-phosphate dehydrogenase committed to the first step of the pentose phosphate pathway, ranked high in fluorescence at position 358 out of 4,596 in the KO screen (**Fig. 4, Suppl. Table S2**). Deletion of *ZWF1* in combination with overexpression of *TKL1*, encoding transketolase 1 responsible for the conversion of xylulose-5-phosphate and ribose-5-phosphate to sedoheptulose-7-phosphate and glyceraldehyde-3-phosphate, is a modification recurrently

employed to rewire the flux of the pentose phosphate pathway (PPP) towards its non-oxidative branch (Brückner et al., 2018; Curran et al., 2013b). This increases the availability of E4P, a precursor known to be rate limiting for the shikimate pathway and CCM production (Horwitz et al., 2015; Suástegui et al., 2016). Since *TKL1* is encoded in the CCM cassette delivered to the libraries **(Fig. 1a)**, the high rank of the *ZWF1* deletion strain corroborates previous strategies for improving CCM synthesis.

Likewise, we found *PAD1*, encoding phenylacrylic acid decarboxylase 1, to rank at position 231 out of 4,983 in the OEx screen (**Fig. 4, Suppl. Table S2**). Pad1 catalyses the prenylation of riboflavin, that is the same reaction catalysed by the cassette-encoded AroY-B (Weber et al., 2017). Prenylated riboflavin is a necessary cofactor in the decarboxylation of PCA to catechol catalysed by AroY.Ciso **(Fig. 1a)**. AroY.Ciso has been shown to be rate limiting in the synthesis of CCM in yeast (Brückner et al., 2018; Curran et al., 2013b; Leavitt et al., 2017; Payer et al., 2017; Weber et al., 2012). The prenylation catalysed by Pad1 is slow (k_{cat} = 12.2 h^{-1}), yet more effective than AroY-B, and co-expression of Pad1 with AroY.C is known to improve availability of prenylated riboflavin (Arunrattanamook and Marsh, 2018; Weber et al., 2017). Altogether, the high rank of *PAD1* in the OEx screen supports the current understanding that insufficient prenylated riboflavin is responsible for the AroY.C bottleneck in CCM synthesis.

Network analysis identifies novel metabolic engineering targets.

We next sought to obtain a broader system-level knowledge of our data in search of novel cellular functions interacting with the synthesis of CCM. We envisioned that gene:gene interactions could be employed to identify clusters of functionally related genes interacting with CCM synthesis. To visualise these interactions, we built networks for the 250 highest- and lowest-ranking gene groups from the YKO and OEx library screens based on associations stored in the STRING database (Szklarczyk et al., 2021)(**Suppl. Fig. S4**). In each of the 250 hit groups, we observed a subset of genes forming densely connected clusters, mostly through ‘database’ and ‘coexpression’ interactions, among a larger number of unconnected genes (**Suppl. Fig. S4**). This suggested that functional gene clusters were indeed present in each hit group.

We isolated these with a k-cliques detection algorithm (Palla et al., 2005). This approach is used to find all cliques (i.e. complete subgraphs) connected to each other by k-1 connections. We set k=3 and a threshold of number of nodes ≥3 to identify cliques with at least three nodes in which all nodes were connected to the rest of the clique with at least k-1=2 interactions. Next, we labelled these cliques, by running Gene Ontology Enrichment Analysis (GOEA) for the gene sets included in them. As a control we also ran this analysis for a group made of 250 randomly selected genes across the dataset.

With this workflow (**Fig. 4a-b**) we found that all hit groups produced at least ten cliques (**Suppl. Fig. S5**). Besides the lowest-ranking 250 group from the YKO library screen, which produced nine cliques with five or more genes, most cliques in all other groups had only three genes. GOEA labelling of the cliques indicated that they represented a wide variety of cellular functions, including multivesicular body sorting and rRNA processing (**Suppl. Fig. S5**). We also note that the control group of randomly chosen genes also produced cliques with GOEA labels (**Suppl. Fig. S5e**). This advised caution as cliques, i.e. networks of functionally related genes, might simply be found by chance when randomly drawing a set of 250 genes from the *S. cerevisiae* gene pool. We therefore carefully inspected the cliques in the hit groups in search of mechanistically meaningful interactions with CCM synthesis.

Firstly, we noticed the presence of a clique enriching the GO term “riboflavin biosynthetic process” composed of genes encoding Fmn1, Rib5 and Rib3, all taking part in riboflavin synthesis (**Fig. 4c**). As mentioned earlier, the high ranking of *PAD1* in the OEx screen, suggests that the supply of the prenylated riboflavin cofactor is limiting. Here, the high-ranking of *FMN1*, *RIB5* and *RIB3* OEx hits suggests that, in addition to its prenylation, insufficient supply of riboflavin might also contribute to the AroY.C decarboxylation bottleneck (**Fig. 4d**).

Secondly, we found three cliques involved in cellular redox homeostasis. This included two cliques taking part in the process of sulphur assimilation for *de novo* methionine synthesis which is the cell’s most important electron sink (Rashida and Laxman, 2021; Thomas and Surdin-Kerjan, 1997). Interestingly, one of these two cliques was found in the top KO group (containing *MET1,3,8,10,14,16,17,28, HIS3* and *ADE*2) while the other in the top-ranking OEx group (containing *MET5,6* and *YILL058W*, **Fig. 4c**). The third clique, linking *SOD1*, *TRX2* and *TSA1* was labelled with ‘antioxidant activity’, and was extracted from the top-ranking hits of the KO screen. The Sod1-chaperone Ccs1, although not included in the clique, was also the strongest hit in the KO screen (**Suppl. Table S2)**. Last, we also noticed that the synthesis and recycling of glutathione, directly downstream of methionine synthesis, formed a hit hotspot, although it didn’t result in the identification of a clique.

Although less direct, the interaction between redox homeostasis and the CCM pathway becomes clear when recognising that E4P availability, CCM’s rate limiting precursor (Suástegui et al., 2016), increases upon shutdown of the oxidative branch of the PPP (Curran et al., 2013b). The oxidative PPP is the cell’s most important source of NADPH and its up-regulation is critical to maintain the NADPH/NADP ratio (Montllor-Albalate et al., 2022; Ralser et al., 2007). Hence, decreased NADPH consumption in mutants unable to assimilate sulphur should inhibit the oxidative PPP and improve E4P availability. Conspicuously, a recent study proposed that Sod1, a superoxide dismutase, had a crucial role in upregulating

the oxidative PPP in response to oxidative stress. This study was found that under oxidative stress, conversion of superoxide into H_2O_2 by Sod1 directly inhibits the *TDH3*-encoded glycolytic enzyme Gapdh which redirects carbon flux into the oxidative PPP (Montllor-Albalate et al., 2022). Conversely, loss of Sod1 (or its chaperone Ccs1) should derepress *TDH3* regardless of the redox state, inhibiting the oxidative PPP. Corroborating this hypothesis, we also found the *TDH3* OEx strain to be a top-100 hit (**Suppl. Table S1)**

Lastly, disruption of the pyruvate dehydrogenase (PDH) complex, the entry point of pyruvate in the TCA cycle, was also found to be a KO hotspot. Two three-genes cliques and a number of additional genes were found at the top of the KO screen (**Fig. 4c**). These included *PDA1* and *LPD1*, which are parts of the PDH complex subunits E1 and E3, respectively **(Fig. 4d)**. Whatsmore, several genes encoding enzymes taking part in the synthesis of the cofactor lipoic acid (*HTD2, MCT1, LIP5)* and its loading (*AIM22, LIP2*) onto PDH subunit E2 were also found to be KO top hits.

In total, network analysis identified multiple gene cliques (**Suppl. Fig. S5**) and within six of these, the mechanistic interactions with CCM biosynthesis could be directly interpreted (**Fig. 4c**). Notably, we are the first to report the interaction between redox homeostasis and CCM synthesis, presumably through a common interaction with the oxidative PPP. Similarly, the network analysis suggested that improving riboflavin cofactor synthesis or introducing a pyruvate bottleneck by impairing the PDH complex should benefit CCM biosynthesis, corroborating previous observations (Blank et al., 2005; Pyne et al., 2018; Weber et al., 2017).

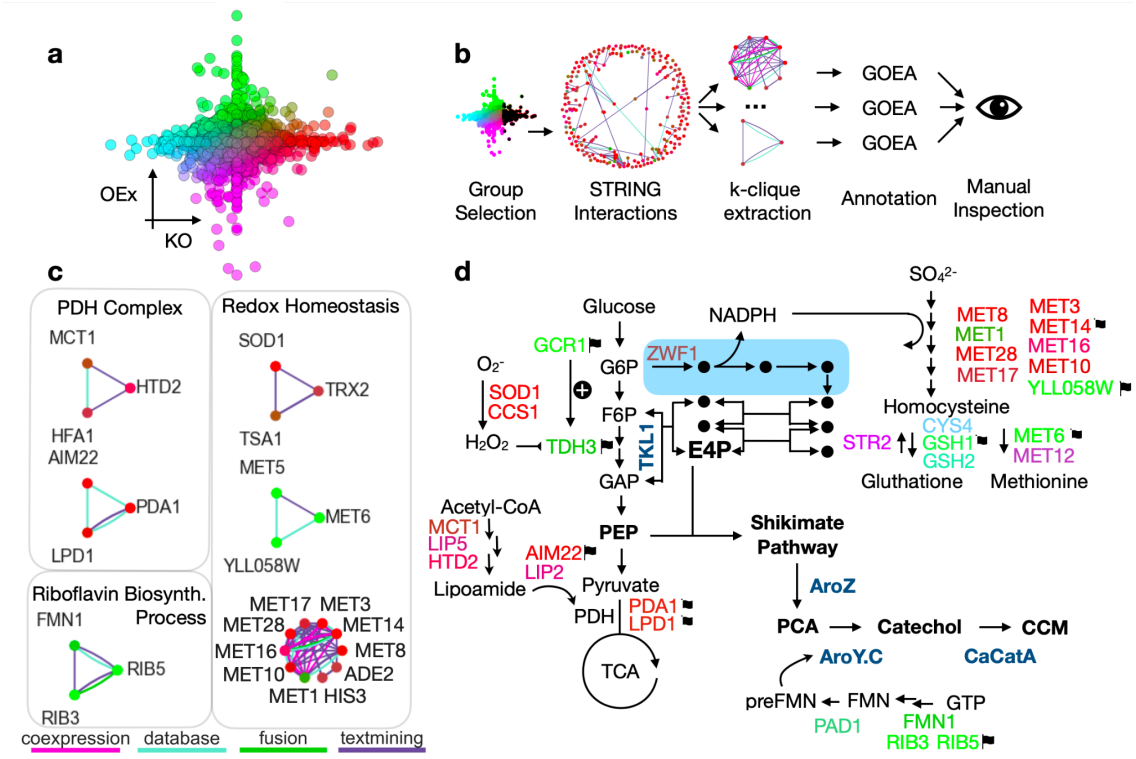


Fig. 4. New cellular functions interacting with the synthesis of CCM. a) KO-OEx colour map. Each point is the mean fluorescence of a gene in the KO screen (x-axis) and the OEx screen (y-axis). Each gene was colour-coded according to its x-y coordinate to map the effects of its KO and OEx on CCM synthesis as shown in panels b-d. **b)** Overview of the hit gene cliques extraction and annotation workflow. Group selection: a group of hits is selected from the screen data. For e.g. KOs causing the strongest fluorescence (black dots). STRING interactions: A STRING interaction network is drawn for the selected group. k-clique extraction: cliques of genes linked by STRING interactions are isolated with the k-clique extraction algorithm. Annotation: GOEA is performed for the set of genes present in the extracted cliques to find enriched GO terms. Manual inspection: the enriched GO terms are inspected for cellular functions interacting with CCM synthesis. **c)** Selected cliques extracted in b) showing mechanistically interpretable interactions with CCM synthesis. Node colours are determined by the colour map in a); the type of STRING interaction is indicated by the bottom legend. **d)** Mapping of the relationship of gene hits found in cliques and discussed in text with CCM synthesis. Gene name colours are determined by colour map in a), genes selected for HPLC validation are indicated with a flag, genes encoded by the CCM cassette are shown in bold blue. CCM intermediates and precursors are shown in bold black.

Experimental validation of host:pathway interaction hits.

In the previous section, CRI-SPA unveiled six gene cliques, which appear to interact with CCM synthesis. As a next step, we decided to validate hits sampled from these cliques (**Fig. 4d**) as well as a few hits from the cellular function category whose interaction with the pathway was less interpretable. We picked eight and six top hits in the OEx and KO screens, respectively. We also picked three low gene hits from the KO screen to further assess the correlation between CRI-SPA fluorescence score and CCM titers (**Suppl. Table S3).**

For the KO hits, we purified strains from screening plates. The integrity of the pathway was verified by genotyping and the sequence of the unique barcodes identifying each gene deletion was verified by sequencing (Giaever et al., 2002). For the OEx hits, we placed each CDS under control of the strong *TEF2* promoter (Sun et al., 2012) and integrated the construct together with the CCM cassette into expression site XII-5 (Jessop-Fabre et al., 2016) of the parental BY4741 strain. To minimise the risk that the hit scores were the result of screening artefacts (e.g. plate, neighbour, agar effects, CCM diffusion in agar) we re-arrayed our validation strains on agar in 384 format interspersed with the control strain BY2. The fluorescence signal obtained in these conditions was well correlated with and less noisy than the original CRI-SPA score (PCC=0.70, p<10e-3, **Fig. 5a**). Again, we observed that the KO strains were generally less fluorescent than the OEx strains and that four of them (*LPD1, PDA1, AIM22, MET14*) were even less fluorescent than the WT. This simple rearraying could constitute a biosensor-mediated intermediate strain selection step, bridging high-throughput CRI-SPA (strains > 10,000s) to lower throughput HPLC (strains < 10s).

To verify that CRI-SPA fluorescence was representative of CCM accumulation, the bottom and top hits from the KO screen were cultivated in liquid Synthetic Complete (SC) medium for 72 h and CCM levels from individual supernatants were measured by HPLC. Here, we found a high correlation (PCC = >0.88) between the CRI-SPA fluorescence score and absolute HPLC measurements of CCM titers across multiple trials despite the batch-to-batch variation in the average CCM titers in between experiments **(Suppl. Fig. S6)**. This robust agreement between CRI-SPA score and CCM titers is a notable achievement given that CRI-SPA measures CCM levels with the help of a biosensor from colonies growing on solid medium.

To assess the ability of our screen to identify high CCM producers, we cultivated both OEx and KO top hits and measured their CCM levels. Importantly, we found that ten out of the fourteen tested top hits showed a significant increase in CCM titers compared to the BY2 reference strain (**Fig. 5b)**. Remarkably, all the selected OEx hits produced significantly more CCM than the reference strain. We also observed that OEx hits seemed to perform better than KO hits. This is in agreement with the observation that the OEx strains were more fluorescent than the KO strains (**Fig. 5a, Suppl. Fig. S3**) and that the difference between top

hits and WT was larger in the OEx screen than in the KO screen (**Fig. 3**). The top four producers were all OEx hits with an improvement in CCM synthesis of 189-286%. The functions of the hits represented a variety of functional areas, from methionine (*YLL058W, MET6, MET14*), to arginine biosynthesis (*ARG5,6*) and the cryptic replication fork blocking gene (*FOB1*, **Suppl. Table S3**). Out of the four non-significant hits, three (*AIM22, LPD1* and *PDA1*) are involved in PDH function. This is a surprise given the number of top hits linked to PDH function found in the KO screen (**Fig. 4c**). We note however that these false positives could have been detected and removed from our analysis by the intermediate re-array step (**Fig. 5a**). Likewise, a biological replicate of the rearrayed OEx screen could not confirm the improved CCM titers of OEx top hits *RIB5* and *GSH1* (**Fig 5b, Suppl. Fig. S7**)

In summary, these validation results demonstrate the capacity of our CRI-SPA screen and data analysis to reliably identify new and diverse gene targets which positive interaction can be used in a metabolic engineering task, such as CCM biosynthesis in yeast cell factories.

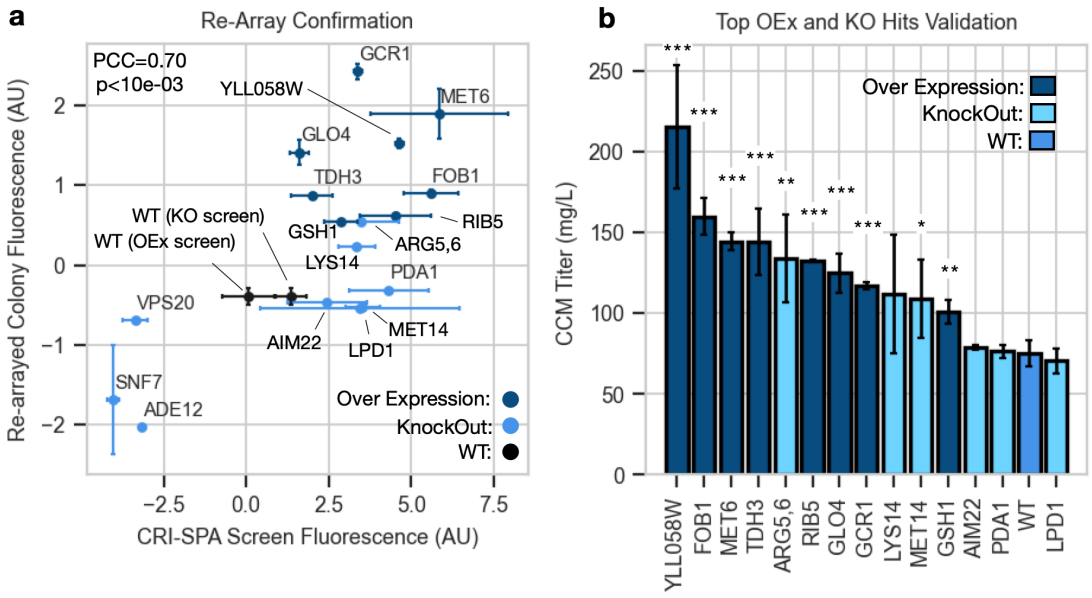


Fig. 5. CRI-SPA hit validation. a) Scatter plot of fluorescences measured in the CRI-SPA screen and from plates of re-arrayed confirmation strains (PCC=0.70, p<10e-3). For the y-axis, the fluorescence of the re-arrayed plates were centred on the mean of their WT and scaled by the plate's standard deviation. Points show the mean of four biological replicates and the error bars show their standard deviations. AU: arbitrary units. **b**) CCM titers for OEx top and bottom hits cultivated in liquid culture. Each bar represents the mean of 3-5 technical replicates. Biological replicate experiment for OEx hit is shown in **Suppl. Fig. S7**. Significant

difference from WT is assessed with a two-sided t-test and indicated by: *<0.05, **<0.01, ***<0.001.

Discussion

The importance of host:pathway interactions has long been recognised (Cardinale and Arkin, 2012; de Lorenzo, 2011) and has motivated the development of a “host-aware” concept in synthetic biology (Boo et al., 2019). Focusing on the expression constructs themselves, successful strategies have sought to isolate them (Darlington et al., 2018; Lou et al., 2012; Meyer et al., 2015) or to minimise the burden they create on their host (Barajas et al., 2022; Ceroni et al., 2018). The importance of the host and its impact on the behaviour of engineered sequences is also increasingly being recognised (Calero and Nikel, 2019; Chan et al., 2023; Khan et al., 2020; Tas et al., 2021). For example, different host strains can show up to ten-fold differences in product titers (Strucko et al., 2015). Adapting the host to its construct opens a high-dimensional design space holding numerous exploitable phenotype maxima. This urges synthetic biologists to find the means to rapidly access and exploit these maxima.

ALE and genome-wide screens arguably search the host space for modifications improving a phenotype of interest. ALE screens the host genome for serendipitous edits and relies on fitness improvement to select the best ones (Mavrommati et al., 2022; Sandberg et al., 2019). Genome-wide screens introduce genetic perturbations, generally with CRISPR, recombineering or transposable elements (Cain et al., 2020; McCarty et al., 2020; Yilmaz et al., 2022), and screen and sort the edited library using massively parallelized reporter assays and next-generation sequencing to identify enriched edits (Bock et al., 2022; Cain et al., 2020). While both ALE and genome-wide screens have been extremely powerful at generating improved strains, the mechanistic learnings they generate are limited. For ALE, extracting the subset of beneficial mutations responsible for the trait improvement requires testing of individual and combinations of edits (Mundhada et al., 2017; Wang et al., 2020). Likewise, for genome-wide screens, a handful of best-performing strains usually dominate the top pool potentially obscuring other beneficial host:pathway interactions (Li et al., 2015; Savitskaya et al., 2019).

By editing each strain of an arrayed genome-wide library, CRI-SPA offers unprecedented functional tractability. The host space is systematically queried along every gene dimension for improvement of a desirable trait. The resulting dataset supports the building of hypotheses on the nature of the functional groups of hits identified. To facilitate this process, we propose in this study a computational strategy using STRING interaction networks, k-clique extraction, and GOEA annotation to identify functional gene clusters

interacting with the synthesis of CCM, or any other analyte for which its quantification can be approximated by image analysis. This workflow has enabled the retrieval of a number of new gene clusters whose interactions could be deciphered as direct (e.g. riboflavin cofactor biosynthesis) and indirect (e.g. redox homeostasis effects on E4P precursor). We show that the top hits selected with the help of this *a priori* knowledge had a high true positive hit rate (10/14 >70%).

In its previous implementation, CRI-SPA relied on the yellowness of the plant pigment betaxanthin to measure strain productivity at high throughput (Cachera et al., 2023). This limited its application to a reduced number of analytes which are directly measurable via their colorimetric (e.g. beta-carotenes) or absorbance properties (e.g. *p*-coumaric acid), and which are frequently employed as optimisation targets in high-throughput screens (Zeng et al., 2020). In this work, the 19kb cassette encoded both the synthesis of CCM and a transcription factor-based biosensor (i.e. BenM MP17_D08) which, without further optimisation, allowed for the read-out of fluorescence directly from high-density agar arrays by image analysis. In the short-term future, we anticipate that the ease of biosensing, as implemented in this study, will open CRI-SPA's applicability to a much wider range of small molecules and biologics which do not create a directly measurable phenotype (Dekker and Polizzi, 2017; Koch et al., 2019), and potentially the application of the CRI-SPA interaction screen as a general learning platform identifying debottlenecking strategies for genetically less tractable microorganisms.

Materials and Methods

Strains and Media

E. coli strains were cultivated in Luria-Bertani medium at 37°C containing 100 mg/L ampicillin to maintain plasmids. Yeast strains were propagated in YPD (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose) medium at 30°C. When appropriate, selection was performed with Nourseothricin (NTC) at 50 mg/L, Hygromycin at 100 mg/L and/or G418 at 100 mg/L in liquid medium and double these concentrations on solid medium.Ura3 counter selection was performed on solid SC medium supplemented with 1 g/L 5-fluoroorotic acid (5-FOA) and 30 mg/L uracil.

The YKO library was purchased from Transomic and the OEx TEF2pr-mCherry library was a kind gift from the Schuldiner lab (Weill et al., 2018). Strains and plasmids used in this study are available in **Suppl. Tables S4-S5**, respectively.

General Cloning Procedure

All DNA sequences destined to cloning were amplified with Phusion U Hotstart (Thermofisher) using primers with USER compatible overhangs, assembled using USER cloning (Nour-Eldin et al., 2006) and transformed by heat shock in *E.coli* Dh5alpha. All plasmids inserts were sequenced and confirmed prior to yeast transformation. Yeast was transformed using the lithium acetate protocol (Gietz and Schiestl, 2007) and correct insertion was verified by PCR validation (DreamTaq Green, Thermo) both upstream and downstream of the integration site.

Refactoring the CCM pathway

Repetitive promoters and terminators used to express the nine CDSs in the CCM pathway were replaced through three rounds of DNA assembly (**Suppl. Fig. S1**).The original promoter for BenM expression (REV1p) and the engineered pCYC_BenO driving yeGFP expression were left unmodified. In the first round, CDSs, promoters and terminators were fused into expression cassettes by USER assembly. This gave plasmids pCCM001-11. In the second round, multiple cassettes were fused and inserted into homology backbones containing flanking homology sequences. The homology sequences on the homology backbones were designed to insert the biosensor and the pathway together or on their own at site chromosomal site XII-5: pCCM021(XII-5Up:clone:HomoB), pCCM022 (HomoB:clone:XII-5Down), pCCM023 (HomoA:clone:HomoB) and pCCM024 (XII-5Up:clone:HomoA). Round two gave plasmids: pCCM012-16,18,20. In the third and final round, the joined expression cassettes were excised from pCCM012-16,18,20 and transformed and assembled into *S. cerevisiae* by homologous recombination. The USER primers used to PCR and fuse all CDSs, promoters, and terminators were automatically generated with our python script (https://github.com/pc2912/CRI-SPA_CCM).

Gene knockout

The up and down homology sequences of the relevant ORF were fused to the kanMX cassette by USER cloning. USER primers were used to generate PCR fragments of the up- and downstream sequences of the ORFs using genomic DNA from BY4741 as template. The relevant PCR fragments were USER-fused to the kanMX cassette in *E. coli* and inserted into the USER cloning site of pCCM023, a modified version plasmid pCfB2909 lacking XII5-up and XII5-down homology sequences. The Up::kanMX::Dw gene-targeting substrates were released from purified plasmids by digestion with NotI prior to transformation into the relevant yeast strain which was selected on YPD containing G418.

HPLC

Standards were prepared by diluting CCM (Sigma-Aldrich; 15992-5G-F) into deionized water to 200 mg/L and stirring until CCM was visibly dissolved. This dilution was heated at 95°C for 1 h and back diluted to obtain 0, 12.5, 25, 50, 100, and 200 mg/L dilutions which were used as standards for the calibration curve.

Strains were plated on selective SC agar medium (+NTC) to select for the CCM pathway. Individual colonies were picked in 2 mL of SC medium without antibiotic and grown overnight. The OD for each strain was adjusted to OD600 = 0.1 in 1 mL of the same medium in deep-well plates in experimental triplicates and grown for 72 h. After 72 h, the deep-well plate was spun at 3,000 *g* for 5 min. Next, 200 uL of supernatant was transferred to a 96-well plate which was tightly sealed and heated at 95°C for 1 h to kill any remaining cells. The plate was spun again at 3,000 *g* for 5 min and a ¼ to ⅛ dilution of the supernatant was used for HPLC quantification. Analysis was run on an Aminex HPX-87H ion exclusion column kept at 60 C using 1 mM H_2SO_4 as solvent injected at 0.6 ml/min. CCM was detected with a Thermo Fisher UV detector at 250 nm.

Flow cytometry analysis

Single colonies were picked in 2 mL SC medium and grown overnight. Overnight cultures were diluted 50 times in 1 mL SC medium in 96-well plates and incubated in a shaking incubator set to 250 rpm, 30°C for 72 h. For CCM induction, SC medium was prepared containing CCM concentrations 710. 355, 177, 88, 8 and 0.8 mg/L. After cultivation, 1:4 water dilution of the cell cultures were analysed in a Quanteon flow cytometer (Novocyte). GFP fluorescence was measured using the B525 channel (emission: 488 nm, absorption 525/45 nm) for 12,000 events with size gate FSC-H > 3000. Flow cytometry data was manually gated and analysed with Python package Cytoflow (Teague, 2022).

CRI-SPA pinning procedure

The pinning procedure was carried out with the Rotor (Singer Instruments) as previously described (Cachera et al., 2023). Briefly, the donor strain was grown overnight in liquid YPD with hygromycin to maintain the CRI-SPA vector to an OD between 0.6 and 2. Cells were washed twice with sterile deionised water to remove the antibiotic. A volume of 500 uL of the washed suspension was spread on YPD Singer Plus Plates which were left to dry in a laminar airflow cabinet until visibly dry. The library (in 384 format) was pinned in 4x4 quadruplicate (1,536 format) on top of the donor strain and incubated for 16 h at 30°C for mating. In our experiments we found that optimal mating efficiency was obtained when little donor library biomass was transferred to the mating plate. If the 384 library plates have big colonies, some colony biomass should be removed by a first uptake with 384 pads.

The mated arrays were replica pinned on the following agar substrates: YP + raffinose + hygromycin + antibiotic1 (24-48 h) > YP + galactose+ hygromycin + antibiotic1 (24 h) > SC 5-FOA (24-48 h) > SC + glucose + NTC + G418. Antibiotic1 is the library-specific antibiotic: G418 for YKO, and NTC for the OEx library. Glucose, raffinose and galactose were supplied at 20 g/L. All incubations were performed at 30°C. Images were acquired from the final SC + glucose + NTC + G418 plate after 48 h of incubation.

Image acquisition and processing.

For fluorescence measurements, screen plates were imaged with a Fusion-FX6 imager (Vilber) using the C480 light source, a F-535 Y2 filter and aperture option of 4 and an exposure time of 240 ms. Because lighting was not homogeneous within the imager, each plate was imaged with a rotation of 0, 90, 180, and 270 degrees. Colony fluorescence was then defined as the mean fluorescence extracted from the four pictures. For colony size, plates were imaged in white light with bottom lighting in a Phenobooth Imager (Singer instruments). Both raw colony size and fluorescence were extracted from plate pictures with the functions of the Pyphe Python package (Kamrad et al., 2020) as performed previously (Cachera et al., 2023).

Data processing and normalisation

The experimental variance affecting colony fluorescence and size, was corrected using the same strategy developed for Synthetic Gene Array developed by the Boone and Andrews lab (Baryshnikova et al., 2010; Wagih et al., 2013). More specifically, the functions: Plate normalisation (N1), Spatial normalisation (N2), Row column effect (N3) and Jackknife filter (F4) were translated from R from (Wagih et al., 2013) and incorporated in our Python workflow. Raw colony fluorescence and colony size were processed in the following order: N1, N2, N3, F4, N1. After this processing, genes with less than three colony replicates were removed from our analysis. Scripts used for the analysis are available on our github repository (https://github.com/pc2912/CRI-SPA_CCM).

Combining data from independent replicate screens

The raw data of two independent screens was passed through the following functions: N1, N2, N3, F4. At this step, N1 was applied so that the median of all plates from two screens were set to be equal. The data of the two screens was then pooled and genes with less than three colony replicates across the two screens were removed from our analysis.

String database and networks

The STRING data was downloaded as file fi4932.protein.links.full.v11.5 from the database website (Szklarczyk et al., 2021). A group of genes was selected based on their ranking in the KO or OEx screens, for example the brightest 250 genes in the KO screen. An undirected graph was built representing genes as nodes and high confidence (confidence > 0.7) STRING associations as edges. Edges were coloured based on the type of association. For node colours, the data of the KO and OEx screens were merged in one dataframe to create dosage scores. If a gene readout was present only in one dataset, its score in the other dataset was assumed to be 0. A heuristic function was created to map the dosage score to an RGB value to maximise visible colour difference between genes on KO vs OEx plot (**Suppl. Fig. S4**).

Gene clusters were extracted by running NetworX's k-clique communities function with k=3 (Palla et al., 2005). GOEA was run for the gene sets contained in each clusters with GOATOOLS (Consortium, 2006; Klopfenstein et al., 2018).

KO Strains recovery and sequencing

For validation of the KO hits, strains were picked from the 1,536 screening plates and restreaked on NTC YPD plates. Fluorescence of the restreaks was verified and a representative colony for each hit was restreaked on a new NTC YPD plate. The integrity of the CCM-cassette was verified by amplifying its upstream and downstream integration sites. To validate the identity of the KO strain, the upstream and downstream barcode unique to each KO strain was amplified with primer pairs U1 + BTX.55 and U2+CCM.124 and Sanger sequenced. Both up and down barcodes were recovered for KO: *PDA1, ARG5,6, LYS14, MET14, LPD1, SNF7, ADE12* and *VPS20.* For *AIM22,* we only recovered the downstream barcode. Strains were cryostocked and used for further fluorescence and HPLC validations.

OEx strains reverse engineering

The assembled CCM cassette was recovered by PCR from BY2 into two parts which were inserted into pCCM021 and pCCM039, resulting in plasmids pCCM042 and pCCM043, respectively. The ORFs of the selected OEx hits: *GCR1, MET6, GLO4, GSH1, TDH3, FOB1, RIB5* and *YLL058W* were amplified with their terminators from BY4741 genome and cloned downstream of TEF2p in pCCM40, resulting in plasmids pCCM055-59,61-63 (**Suppl. Table S5**). The insertion sites of pCCM021, pCCM039, and pCCM040 are flanked with matching homology sequences allowing for integration of their inserts into yeast genomic site XII-5 by homologous recombination. To create the OEx hits strain, the inserts of pCCM042 and pCCM043 and of either of pCCM055-59,61-63 were transformed together with pCF3050 in BY4741 harbouring Cas9 encoded on pHO38 and were selected on YPD + hygromycin + NTC + G418.

Rearraying on 384 arrays

Because the OEx top hits were more fluorescent than the KO top hits, if arrayed on the same agar plate, the KO top hits' fluorescence was "blinded" by that of the OEx top hits. To avoid this, we rearrayed the OEx hits and KO hits on two different agar arrays. The hit strains were pinned in square quadruplicates and interspersed with control strain BY2. BY2 was also placed at the edges of the array. In this way, all hit strains had BY2 as a direct neighbour and were not exposed to edge effects. Because of the high density of hits on the plates the "Boone and Andrews lab" correction (Wagih et al., 2013) was not applied. Rather, each plate data was standardised by subtracting the BY2 means and dividing by the plate's standard deviation.

Code and Data availability

The scripts used to design primers, extract the data from raw images, conduct analyses and reproduce the figures shown in this study are available on our Github repository: https://github.com/pc2912/CRI-SPA_CCM. The fluorescent and white light images for the OEx and one repeat of the KO screens are available to run the extraction and correction analyses. The raw and corrected fluorescence and colony size data for all screens mentioned in this study KO and OEx are also available on this repository.

Acknowledgements

This study is supported by grants from the Novo Nordisk Foundation (NNF20CC0035580 and NNF19SA0035438) to MKJ and PP-YJC. We thank Prof. Lars Jul Jensen for his feedback on the progress of this work and his advice on using the STRING database. We thank Prof. Maya Schuldiner for kindly sending us the TEF2p SWAT overexpression library.

Author contributions

P.C., U.H.M. and M.K.J. designed research. P.C., N.C.K and A.R. performed research. P.C. analyzed the data. T.S, U.H.M and M.K.J, supervised the research. P.C. and M.K.J. wrote the paper.

Conflict of Interest

The authors declare no financial interest.

References

Arunrattanamook, N., Marsh, E.N.G., 2018. Kinetic characterization of prenyl-flavin synthase from Saccharomyces cerevisiae. Biochemistry 57, 696–700.

Bao, Z., HamediRad, M., Xue, P., Xiao, H., Tasan, I., Chao, R., Liang, J., Zhao, H., 2018. Genome-scale engineering of Saccharomyces cerevisiae with single-nucleotide precision. Nat. Biotechnol. 36, 505–508.

Barajas, C., Huang, H.-H., Gibson, J., Sandoval, L., Del Vecchio, D., 2022. Feedforward growth rate control mitigates gene activation burden. Nat. Commun. 13, 7054.

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S., others, 2010. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat. Methods 7, 1017–1024.

Blank, L.M., Kuepfer, L., Sauer, U., 2005. Large-scale 13 C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. Genome Biol. 6, 1–16.

Bock, C., Datlinger, P., Chardon, F., Coelho, M.A., Dong, M.B., Lawson, K.A., Lu, T., Maroc, L., Norman, T.M., Song, B., others, 2022. High-content CRISPR screening. Nat. Rev. Methods Primer 2, 8.

Boo, A., Ellis, T., Stan, G.-B., 2019. Host-aware synthetic biology. Curr. Opin. Syst. Biol. 14, 66–72.

Borkowski, O., Bricio, C., Murgiano, M., Rothschild-Mancinelli, B., Stan, G.-B., Ellis, T., 2018. Cell-free prediction of protein expression costs for growing cells. Nat. Commun. 9, 1457.

Brückner, C., Oreb, M., Kunze, G., Boles, E., Tripp, J., 2018. An expanded enzyme toolbox for production of cis, cis-muconic acid and other shikimate pathway derivatives in Saccharomyces cerevisiae. FEMS Yeast Res. 18, foy017.

Cachera, P., Olsson, H., Coumou, H., Jensen, M.L., Sánchez, B.J., Strucko, T., van den Broek, M., Daran, J.-M., Jensen, M.K., Sonnenschein, N., others, 2023. CRI-SPA: a high-throughput method for systematic genetic editing of yeast libraries. Nucleic Acids Res. gkad656.

Cain, A.K., Barquist, L., Goodman, A.L., Paulsen, I.T., Parkhill, J., van Opijnen, T., 2020. A decade of advances in transposon-insertion sequencing. Nat. Rev. Genet. 21, 526–540.

Calero, P., Nikel, P.I., 2019. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. Microb. Biotechnol. 12, 98–124.

Canton, B., Labno, A., Endy, D., 2008. Refinement and standardization of synthetic biological parts and devices. Nat. Biotechnol. 26, 787–793.

Cardinale, S., Arkin, A.P., 2012. Contextualizing context for synthetic biology–identifying causes of failure of synthetic biological systems. Biotechnol. J. 7, 856–866.

Ceroni, F., Algar, R., Stan, G.-B., Ellis, T., 2015. Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat. Methods 12, 415–418.

Ceroni, F., Boo, A., Furini, S., Gorochowski, T.E., Borkowski, O., Ladak, Y.N., Awan, A.R., Gilbert, C., Stan, G.-B., Ellis, T., 2018. Burden-driven feedback control of gene expression. Nat. Methods 15, 387–393.

Chan, D.T.C., Baldwin, G.S., Bernstein, H.C., 2023. Revealing the chassis-effect on a broad-host-range genetic switch and its concordance with interspecies bacterial physiologies. bioRxiv 2023–02.

Consortium, G.O., 2006. The gene ontology (GO) project in 2006. Nucleic Acids Res. 34, D322–D326.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., others, 2010. The genetic landscape of a cell. science 327, 425–431.

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W.,

Usaj, M., Hanchard, J., Lee, S.D., others, 2016. A global genetic interaction network maps a wiring diagram of cellular function. Science 353, aaf1420.
Costello, A., Badran, A.H., 2021. Synthetic biological circuits within an orthogonal central dogma. Trends Biotechnol. 39, 59–71.
Curran, K.A., Karim, A.S., Gupta, A., Alper, H.S., 2013a. Use of expression-enhancing terminators in Saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications. Metab. Eng. 19, 88–97.
Curran, K.A., Leavitt, J.M., Karim, A.S., Alper, H.S., 2013b. Metabolic engineering of muconic acid production in Saccharomyces cerevisiae. Metab. Eng. 15, 55–66.
d'Aquino, A.E., Kim, D.S., Jewett, M.C., 2018. Engineered ribosomes for basic science and synthetic biology. Annu. Rev. Chem. Biomol. Eng. 9, 311–340.
d'Oelsnitz, S., Love, J.D., Diaz, D.J., Ellington, A.D., 2022. GroovDB: A database of ligand-inducible transcription factors. ACS Synth. Biol. 11, 3534–3537.
Dahl, R.H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T.S., Redding-Johanson, A.M., Petzold, C.J., Mukhopadhyay, A., Lee, T.S., Adams, P.D., others, 2013. Engineering dynamic pathway regulation using stress-response promoters. Nat. Biotechnol. 31, 1039–1046.
Darlington, A.P., Kim, J., Jiménez, J.I., Bates, D.G., 2018. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. Nat. Commun. 9, 695.
de Lorenzo, V., 2011. Beware of metaphors: chasses and orthogonality in synthetic biology. Bioeng. Bugs 2, 3–7.
Dekker, L., Polizzi, K.M., 2017. Sense and sensitivity in bioprocessing—detecting cellular metabolites with biosensors. Curr. Opin. Chem. Biol. 40, 31–36.
Friedland, A.E., Lu, T.K., Wang, X., Shi, D., Church, G., Collins, J.J., 2009. Synthetic gene networks that count. science 324, 1199–1202.
Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., others, 2002. Functional profiling of the Saccharomyces cerevisiae genome. nature 418, 387–391.
Giaever, G., Nislow, C., 2014. The yeast deletion collection: a decade of functional genomics. Genetics 197, 451–465.
Gietz, R.D., Schiestl, R.H., 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2, 31–34.
Gyorgy, A., Jiménez, J.I., Yazbek, J., Huang, H.-H., Chung, H., Weiss, R., Del Vecchio, D., 2015. Isocost lines describe the cellular economy of genetic circuits. Biophys. J. 109, 639–646.
Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W., 1999. From molecular to modular cell biology. Nature 402, C47–C52.
Horwitz, A.A., Walter, J.M., Schubert, M.G., Kung, S.H., Hawkins, K., Platt, D.M., Hernday, A.D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, S.S., others, 2015. Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. Cell Syst. 1, 88–96.
Jensen, E.D., Ambri, F., Bendtsen, M.B., Javanpour, A.A., Liu, C.C., Jensen, M.K., Keasling, J.D., 2021. Integrating continuous hypermutation with high-throughput screening for optimization of cis, cis-muconic acid production in yeast. Microb. Biotechnol. 14, 2617–2626.
Jessop-Fabre, M.M., Jakočiūnas, T., Stovicek, V., Dai, Z., Jensen, M.K., Keasling, J.D., Borodina, I., 2016. EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into Saccharomyces cerevisiae via CRISPR-Cas9. Biotechnol. J. 11, 1110–1117.
Jones, D., Reusser, U., Braus, G., 1991. Molecular cloning, characterization and analysis of the regulation of the AR02 gene, encoding chorismate synthase, of Saccharomyces cerevisiae. Mol. Microbiol. 5, 2143–2153.
Kamrad, S., Rodríguez-López, M., Cotobal, C., Correia-Melo, C., Ralser, M., Bähler, J., 2020. Pyphe, a python toolbox for assessing microbial growth and cell viability in

high-throughput colony screens. Elife 9, e55160.
Khan, N., Yeung, E., Farris, Y., Fansler, S.J., Bernstein, H.C., 2020. A broad-host-range event detector: expanding and quantifying performance between Escherichia coli and Pseudomonas species. Synth. Biol. 5, ysaa002.
Klopfenstein, D., Zhang, L., Pedersen, B.S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C.J., Yunes, J.M., Botvinnik, O., Weigel, M., others, 2018. GOATOOLS: A Python library for Gene Ontology analyses. Sci. Rep. 8, 1–17.
Koch, M., Pandi, A., Borkowski, O., Batista, A.C., Faulon, J.-L., 2019. Custom-made transcriptional biosensors for metabolic engineering. Curr. Opin. Biotechnol. 59, 78–84.
Künzler, M., Paravicini, G., Egli, C.M., Irniger, S., Braus, G.H., 1992. Cloning, primary structure and regulation of the ARO4 gene, encoding the tyrosine-inhibited 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from Saccharomyces cerevisiae. Gene 113, 67–74.
Lazebnik, Y., 2002. Can a biologist fix a radio?—Or, what I learned while studying apoptosis. Cancer Cell 2, 179–182.
Leavitt, J.M., Tong, A., Tong, J., Pattie, J., Alper, H.S., 2016. Coordinated transcription factor and promoter engineering to establish strong expression elements in Saccharomyces cerevisiae. Biotechnol. J. 11, 866–876.
Leavitt, J.M., Wagner, J.M., Tu, C.C., Tong, A., Liu, Y., Alper, H.S., 2017. Biosensor-enabled directed evolution to improve muconic acid production in Saccharomyces cerevisiae. Biotechnol. J. 12, 1600687.
Li, S., Si, T., Wang, M., Zhao, H., 2015. Development of a synthetic malonyl-CoA sensor in Saccharomyces cerevisiae for intracellular metabolite monitoring and genetic screening. ACS Synth. Biol. 4, 1308–1315.
Lian, J., HamediRad, M., Hu, S., Zhao, H., 2017. Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. Nat. Commun. 8, 1688.
Liu, C.C., Jewett, M.C., Chin, J.W., Voigt, C.A., 2018. Toward an orthogonal central dogma. Nat. Chem. Biol. 14, 103–106.
Lou, C., Stanton, B., Chen, Y.-J., Munsky, B., Voigt, C.A., 2012. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. Nat. Biotechnol. 30, 1137–1142.
Mavrommati, M., Daskalaki, A., Papanikolaou, S., Aggelis, G., 2022. Adaptive laboratory evolution principles and applications in industrial biotechnology. Biotechnol. Adv. 54, 107795.
McCarty, N.S., Graham, A.E., Studená, L., Ledesma-Amaro, R., 2020. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. Nat. Commun. 11, 1–13.
Meyer, A.J., Ellefson, J.W., Ellington, A.D., 2015. Directed evolution of a panel of orthogonal T7 RNA polymerase variants for in vivo or in vitro synthetic circuitry. ACS Synth. Biol. 4, 1070–1076.
Montllor-Albalate, C., Kim, H., Thompson, A.E., Jonke, A.P., Torres, M.P., Reddi, A.R., 2022. Sod1 integrates oxygen availability to redox regulate NADPH production and the thiol redoxome. Proc. Natl. Acad. Sci. 119, e2023328119.
Mundhada, H., Seoane, J.M., Schneider, K., Koza, A., Christensen, H.B., Klein, T., Phaneuf, P.V., Herrgard, M., Feist, A.M., Nielsen, A.T., 2017. Increased production of L-serine in Escherichia coli through adaptive laboratory evolution. Metab. Eng. 39, 141–150.
Nielsen, A.A., Der, B.S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E.A., Ross, D., Densmore, D., Voigt, C.A., 2016. Genetic circuit design automation. Science 352, aac7341.
Nour-Eldin, H.H., Hansen, B.G., Nørholm, M.H., Jensen, J.K., Halkier, B.A., 2006. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res. 34, e122–e122.
Palla, G., Derényi, I., Farkas, I., Vicsek, T., 2005. Uncovering the overlapping community structure of complex networks in nature and society. nature 435, 814–818.
Payer, S.E., Marshall, S.A., Bärland, N., Sheng, X., Reiter, T., Dordic, A., Steinkellner, G., Wuensch, C., Kaltwasser, S., Fisher, K., others, 2017. Regioselective

para-carboxylation of catechols with a prenylated flavin dependent decarboxylase. Angew. Chem. Int. Ed. 56, 13893–13897.
Pyne, M.E., Narcross, L., Melgar, M., Kevvai, K., Mookerjee, S., Leite, G.B., Martin, V.J., 2018. An engineered Aro1 protein degradation approach for increased cis, cis-muconic acid biosynthesis in Saccharomyces cerevisiae. Appl. Environ. Microbiol. 84, e01095-18.
Ralser, M., Wamelink, M.M., Kowald, A., Gerisch, B., Heeren, G., Struys, E.A., Klipp, E., Jakobs, C., Breitenbach, M., Lehrach, H., others, 2007. Dynamic rerouting of the carbohydrate flux is key to counteracting oxidative stress. J. Biol. 6, 1–18.
Rashida, Z., Laxman, S., 2021. The pentose phosphate pathway and organization of Metabolic Networks Enabling Growth Programs. Curr. Opin. Syst. Biol. 28, 100390.
Rogers, J.K., Taylor, N.D., Church, G.M., 2016. Biosensor-based engineering of biosynthetic pathways. Curr. Opin. Biotechnol. 42, 84–91.
Sandberg, T.E., Salazar, M.J., Weng, L.L., Palsson, B.O., Feist, A.M., 2019. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. Metab. Eng. 56, 1–16.
Savitskaya, J., Protzko, R.J., Li, F.-Z., Arkin, A.P., Dueber, J.E., 2019. Iterative screening methodology enables isolation of strains with improved properties for a FACS-based screen and increased L-DOPA production. Sci. Rep. 9, 5815.
Segall-Shapiro, T.H., Meyer, A.J., Ellington, A.D., Sontag, E.D., Voigt, C.A., 2014. A ‘resource allocator’for transcription based on a highly fragmented T7 RNA polymerase. Mol. Syst. Biol. 10, 742.
Shopera, T., He, L., Oyetunde, T., Tang, Y.J., Moon, T.S., 2017. Decoupling resource-coupled gene expression in living cells. ACS Synth. Biol. 6, 1596–1604.
Skjoedt, M.L., Snoek, T., Kildegaard, K.R., Arsovska, D., Eichenberger, M., Goedecke, T.J., Rajkumar, A.S., Zhang, J., Kristensen, M., Lehka, B.J., others, 2016. Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. Nat. Chem. Biol. 12, 951–958.
Snoek, T., Chaberski, E.K., Ambri, F., Kol, S., Bjørn, S.P., Pang, B., Barajas, J.F., Welner, D.H., Jensen, M.K., Keasling, J.D., 2020. Evolution-guided engineering of small-molecule biosensors. Nucleic Acids Res. 48, e3–e3.
Snoek, T., Romero-Suarez, D., Zhang, J., Ambri, F., Skjoedt, M.L., Sudarsan, S., Jensen, M.K., Keasling, J.D., 2018. An orthogonal and pH-tunable sensor-selector for muconic acid biosynthesis in yeast. ACS Synth. Biol. 7, 995–1003.
Strucko, T., Magdenoska, O., Mortensen, U.H., 2015. Benchmarking two commonly used Saccharomyces cerevisiae strains for heterologous vanillin-β-glucoside production. Metab. Eng. Commun. 2, 99–108.
Suástegui, M., Guo, W., Feng, X., Shao, Z., 2016. Investigating strain dependency in the production of aromatic compounds in Saccharomyces cerevisiae. Biotechnol. Bioeng. 113, 2676–2685.
Suástegui, M., Ng, C.Y., Chowdhury, A., Sun, W., Cao, M., House, E., Maranas, C.D., Shao, Z., 2017. Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in Saccharomyces cerevisiae for high production of polymer and drug precursors. Metab. Eng. 42, 134–144.
Sun, J., Jeffryes, J.G., Henry, C.S., Bruner, S.D., Hanson, A.D., 2017. Metabolite damage and repair in metabolic engineering design. Metab. Eng. 44, 150–159.
Sun, J., Shao, Z., Zhao, Hua, Nair, N., Wen, F., Xu, J.-H., Zhao, Huimin, 2012. Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in Saccharomyces cerevisiae. Biotechnol. Bioeng. 109, 2082–2092.
Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., others, 2021. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 49, D605–D612.
Tas, H., Grozinger, L., Stoof, R., de Lorenzo, V., Goñi-Moreno, Á., 2021. Contextual dependencies expand the re-usability of genetic inverters. Nat. Commun. 12, 355.

Teague, B., 2022. Cytoflow: A python toolbox for flow cytometry. bioRxiv 2022–07.
Thomas, D., Surdin-Kerjan, Y., 1997. Metabolism of sulfur amino acids in Saccharomyces cerevisiae. Microbiol. Mol. Biol. Rev. 61, 503–532.
Wagih, O., Usaj, M., Baryshnikova, A., VanderSluis, B., Kuzmin, E., Costanzo, M., Myers, C.L., Andrews, B.J., Boone, C.M., Parts, L., 2013. SGAtools: one-stop analysis and visualization of array-based genetic interaction screens. Nucleic Acids Res. 41, W591–W596.
Wang, G., Møller-Hansen, I., Babaei, M., D'Ambrosio, V., Christensen, H.B., Darbani, B., Jensen, M.K., Borodina, I., 2021. Transportome-wide engineering of Saccharomyces cerevisiae. Metab. Eng. 64, 52–63.
Wang, G., Øzmerih, S., Guerreiro, R., Meireles, A.C., Carolas, A., Milne, N., Jensen, M.K., Ferreira, B.S., Borodina, I., 2020. Improvement of cis, cis-muconic acid production in Saccharomyces cerevisiae through biosensor-aided genome engineering. ACS Synth. Biol. 9, 634–646.
Weber, C., Brückner, C., Weinreb, S., Lehr, C., Essl, C., Boles, E., 2012. Biosynthesis of cis, cis-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by Saccharomyces cerevisiae. Appl. Environ. Microbiol. 78, 8421–8430.
Weber, H.E., Gottardi, M., Brückner, C., Oreb, M., Boles, E., Tripp, J., 2017. Requirement of a functional flavin mononucleotide prenyltransferase for the activity of a bacterial decarboxylase in a heterologous muconic acid pathway in Saccharomyces cerevisiae. Appl. Environ. Microbiol. 83, e03472-16.
Weill, U., Yofe, I., Sass, E., Stynen, B., Davidi, D., Natarajan, J., Ben-Menachem, R., Avihou, Z., Goldman, O., Harpaz, N., others, 2018. Genome-wide SWAp-Tag yeast libraries for proteome exploration. Nat. Methods 15, 617–622.
Weiße, A.Y., Oyarzún, D.A., Danos, V., Swain, P.S., 2015. Mechanistic links between cellular trade-offs, gene expression, and growth. Proc. Natl. Acad. Sci. 112, E1038–E1047.
Wood, V., Lock, A., Harris, M.A., Rutherford, K., Bähler, J., Oliver, S.G., 2019. Hidden in plain sight: what remains to be discovered in the eukaryotic proteome? Open Biol. 9, 180241.
Wu, G., Yan, Q., Jones, J.A., Tang, Y.J., Fong, S.S., Koffas, M.A., 2016. Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. Trends Biotechnol. 34, 652–664.
Yilmaz, S., Nyerges, A., van der Oost, J., Church, G.M., Claassens, N.J., 2022. Towards next-generation cell factories by rational genome-scale engineering. Nat. Catal. 5, 751–765.
Yofe, I., Weill, U., Meurer, M., Chuartzman, S., Zalckvar, E., Goldman, O., Ben-Dor, S., Schütze, C., Wiedemann, N., Knop, M., others, 2016. One library to make them all: streamlining the creation of yeast libraries via a SWAp-Tag strategy. Nat. Methods 13, 371–378.
Zeng, W., Guo, L., Xu, S., Chen, J., Zhou, J., 2020. High-throughput screening technology in industrial biotechnology. Trends Biotechnol. 38, 888–906.
Zhang, J., Hansen, L.G., Gudich, O., Viehrig, K., Lassen, L.M., Schrübbers, L., Adhikari, K.B., Rubaszka, P., Carrasquer-Alvarez, E., Chen, L., others, 2022. A microbial supply chain for production of the anti-cancer drug vinblastine. Nature 609, 341–347.
Zhang, J., Petersen, S.D., Radivojevic, T., Ramirez, A., Pérez-Manríquez, A., Abeliuk, E., Sánchez, B.J., Costello, Z., Chen, Y., Fero, M.J., others, 2020. Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. Nat. Commun. 11, 4880.
Zong, Y., Zhang, H.M., Lyu, C., Ji, X., Hou, J., Guo, X., Ouyang, Q., Lou, C., 2017. Insulated transcriptional elements enable precise design of genetic circuits. Nat. Commun. 8, 52.

Chapter 5

Conclusion and Perspectives

5.1 Summary

With the advancement of molecular biology, biologists have become able to read, write, and copy DNA and assemble sequences of their design to program new functions in living (micro-)organisms. This endeavour gave birth to the field of synthetic biology in the early 2000s, which championed the application of traditional engineering principles for the design of biological functions. Despite its elegance, the reductionist conception of living systems oversees the reality of pervasive genetic interactions. In practice, genetic interactions thwart the rational design of synthetic biological functions and their construction still involves a large amount of trial-and-error, i.e. tinkering.

The working objective of this thesis was to propose a new engineering framework that can integrate biological interactions into designs, be broadly applicable and mechanistically tractable.

After an unfruitful attempt at guiding the search for genetic interactions with machine learning presented in Chapter 2, I turned to high throughput screening with CRI-SPA. CRI-SPA systematically screens for interactions between a metabolic feature and every modified host locus present in a yeast library. This breadth gives CRI-SPA high mechanistic tractability as cellular functions interacting with the synthesis of the product can be inferred from the types of genes hits identified. In the example of betaxanthins, a simple GO term Enrichment Analysis (GOEA) (Klopfenstein et al., 2018) on the pool of top gene hits found that mitochondrial disruption consistently improved betaxanthin synthesis. Following this proof of concept presented in Chapter 3, we adapted CRI-SPA to function for the screening of cis-cis-Muconic acid (CCM) via the intermediate reporting of a biosensor, opening CRI-SPA applicability to a much wider range of compounds and cellular processes (Chapter 4).

The work presented in Chapters 3 and 4 proves that mechanistically tractable interactions between a host gene and a heterologous pathway can be exploited to improve the productivity of the wide variety of products which can now be reported with a biosensor. Yet, the search I report in this thesis has been limited to 'first order' interactions, that is, the interactions taking place between a metabolic pathway and one modified host locus. Although the results of these efforts are encouraging, they only scratch the top of the interaction space. The following perspective describes how CRI-SPA could be used to search a much larger interaction space arising from higher order combinations of modified host loci and what to expected from this search.

5.2 Higher order genetic interactions

5.2.1 Adapting CRI-SPA to reach higher order interactions.

From a gene-editing point of view, minor modifications can be made to the current CRI-SPA donor strain to enable higher order interaction screening. In its simplest form, the metabolic pathway would be inserted to disrupt the locus of a gene which was identified as a KO hit in a first screen. A donor strain harbouring this design would systematically deliver the pathway into the locus and dirsupt the gene in the recipient library **(fig. 5.1b)**. Conversely, the best gene identified from a first overexpression screen can be added to the pathway in the donor strain to co-deliver it the next screen **(fig. 5.1c)**. In theory this strategy can be repeated to combine multiple over-expression

loci.

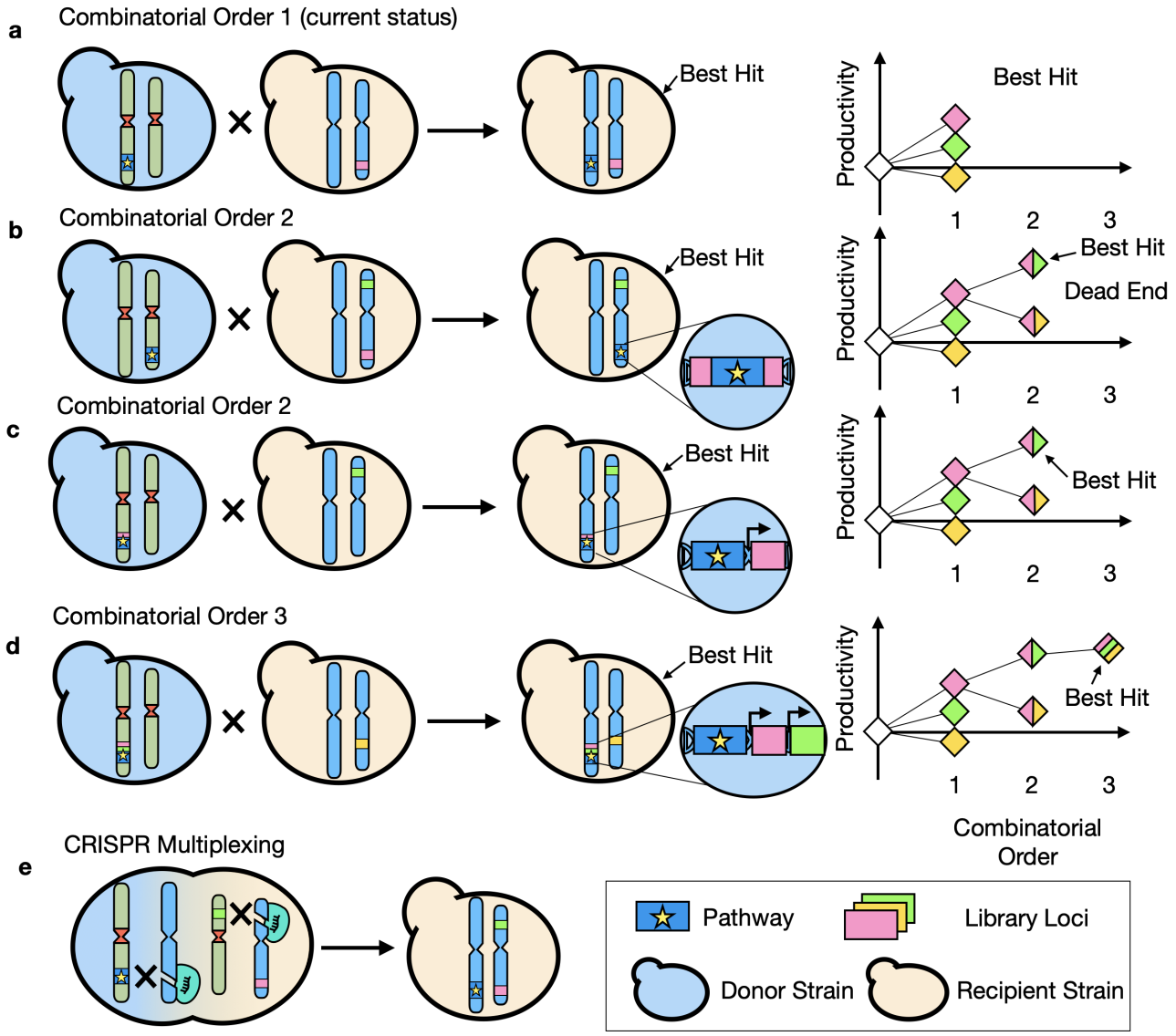


Figure 5.1: **Adapting CRI-SPA to Screen for Higher Order Interactions**
Different strategies can be developed to screen higher order interactions with CRI-SPA. Left shows the cross of the starting donor strain and a library strain generating the final edited strain. Right shows the greedy search iteratively combining the best loci. **a)** Current stage of development. CRI-SPA screens for interactions between a metabolic pathway and one locus for all loci of a library.**b)** The pathway is inserted in a gene locus causing its deletion. This screens for interactions between the pathway, one KO, and one locus for all loci of a library. This strategy is a dead end since no more KO can be introduced in this way. **c)** A gene OEx cassette is added to the pathway. This screens for interactions between the pathway, one OEx, and one locus for all loci of a library. **d)** Same as c) but two OEx cassettes are added to the pathway. **e)** CRI-SPA is developped to support CRISPR multiplexing. Here, multiple edits can be introduced in the library in addition to the delivery of the pathway.

To screen for higher order interactions, the CRI-SPA donor strain could also be adapted to perform CRISPR multiplexing. We showed in Chapter 3 that CRI-SPA can be performed without marker selection for the genetic feature. This opens the possibility to perform multiple edits at once with CRISPR multiplexing which has been highly optimized in the past few years (McCarty et al., 2020; Zhang et al., 2019)

In order to perform multiplexing with CRI-SPA, the individual sgRNAs should be extremely efficient. Indeed, the editing efficiency is known to drop when targeting more than two sites (Ryan et al., 2014; Jakočiūnas, Bonde,

Herrgård, Harrison, Kristensen, Pedersen, Jensen and Keasling, 2015; Jakociunas, Rajkumar, Zhang, Arsovska, Rodriguez, Jendresen, Skjødt, Nielsen, Borodina, Jensen et al., 2015). This editing efficiency would be expected to drop further if the edits have a negative impact on fitness. However, since a CRI-SPA screen would target the same sites in all recipient strain, efficient sgRNA would only need to be designed and verified for those sites. Here, the efficiency could be verified by transferring multiple fluorescent proteins at once into a recipient cell with the same pinning procedure used for the screen. The edited colonies in array format could be analysed with flow cytometry to quantify the percent of edited cells for each edit in each colony.

The CRI-SPA vector would also have to be modified to encode multiple sgRNAs. Here, two sgRNA could be driven by their own promoter. If higher order interactions require three or more sgRNAs, expressing several sgRNAs in one transcript and relying on post-transcriptional processing mechanisms can be considered. For example, sgRNAs expressed as one array can be flanked with RNA cleaving sequences such as the endoribonuclease Csy4 (Ferreira et al., 2019; Lian et al., 2017). The self processing ability of some CRISPR variants can also be leveraged to array several sgRNA within a single expression cassette. For example, the self-sufficient Cas12a allows for sgRNA arraying without the requirement of extra heterologous or host factors (Tak et al., 2017; Zetsche et al., 2017; Liao et al., 2019).

5.2.2 Higher order interactions: what to expect?

If adapted to screening for higher order interactions, CRI-SPA would represent a precious experimental tool for the exploration of the interaction landscape. In this framework, a yeast library would be screened iteratively to find the next best interactions and accumulate them in a producing strain. The implementation of such a systematic search along every host-gene dimension into a directed evolution process would be to my awareness a first in metabolic engineering. Excitingly, the outcome of such a search depends on the topology of the interaction landscape which is far from being completely understood (Bank et al., 2016). For example, burning unknowns regarding the convergence of parallel searches towards the same hits or the appearance of idiosyncrasies (i.e. the type of new hits depends on the current combination of accumulated hits) in the searching process could be investigated. The evolutionary biology literature is rich with theoretical models, both verbal and mathematical, describing the dynamics of evolution within the topology of the multidimensional landscape (Gavrilets, 2010). So what should we expect when entering this high dimensional space?

The conceptualisation of genotype-phenotype relationships as a landscape was proposed by Sewall Wright in 1932 as a framework to describe the evolutionary paths of a set of genes evolving towards a fitness maxima (Wright et al., 1932). According to this metaphor, a sequence is considered a point in a hyper-space where each dimension represents a possible genetic modification. An additional dimension is reserved to hold the phenotype associated with the sequence, generally fitness. By undergoing a mutation, a starting sequence can access a neighbouring state and natural selection pushes each step in a upward walk towards a (local) maximum **(fig. 5.2a)**.

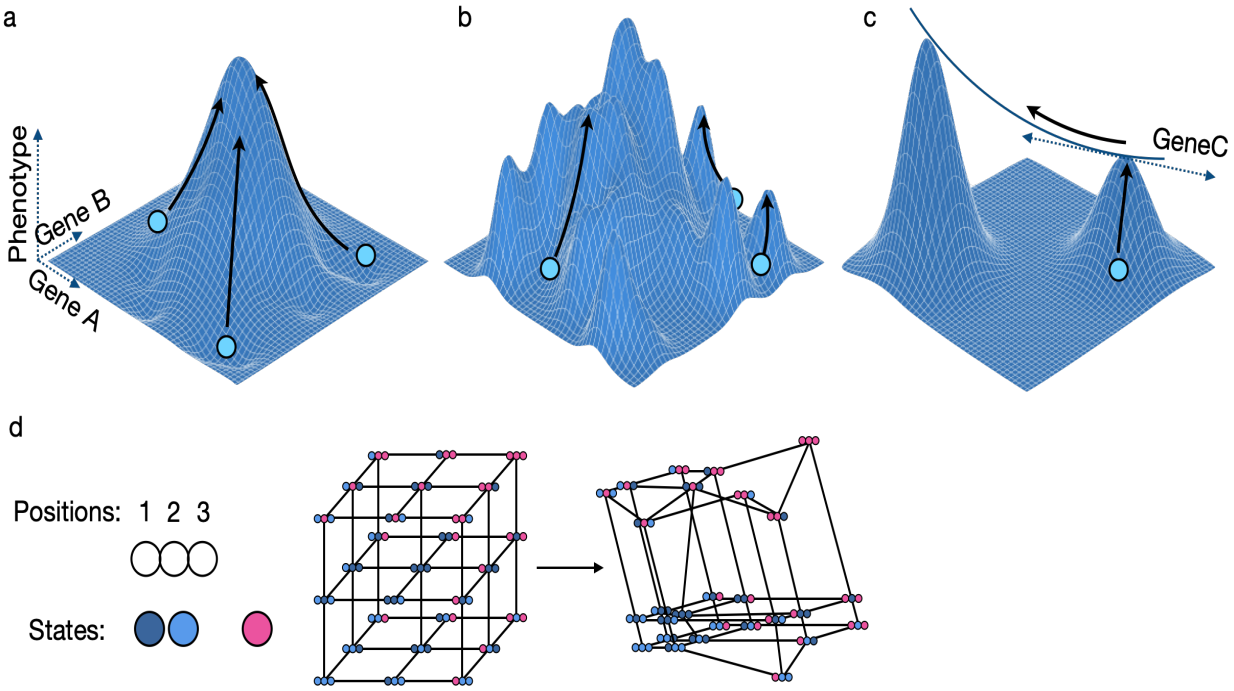


Figure 5.2: **The Phenotype Landscape**

a) In a smooth landscape, all starting sequences converge to the global maximum. **b)** In a rugged landscape, evolving sequences might become trapped in local maxima. **c)** In higher dimensions, peaks are unlikely to be maximas in all dimensions enabling the bridging of low fitness valleys. **d)** Landscape arising from, Left: three positions (e.g. genes, amino acids, etc.) which can adopt three states (e.g. substitution). Here the pink state is depicted as physiologically more different than the light and dark blue states. Center: In the traditional representation of the landscape, positions are orthogonal and states are equidistant. Right: biological systems are likely none-linear and "dimensions" not orthogonal.

Defining the density of local maxima in the landscape (i.e. its "ruggedness") is particularly important to understanding the ability of natural selection to access sequence diversity and its speed of convergence (i.e. the number of steps needed) **(fig. 5.2b)**. Intuitively, higher dimensional spaces should be less prone to local maxima as more neutral dimensions should form saddle points allowing an exit of a maximum **(fig. 5.2c)**. Yet, theoretical biologists has proposed that the number of maxima grows faster than the number of possible paths as the number of dimension increases, resulting in a highly rugged landscape (Kauffman and Levin, 1987). However, the mathematical framework of Kauffman and Levin relies on the assumption that every sequence variant in the space has a distinct fitness rank and that the space is uncorrelated (i.e. the fitness of variants are randomly assigned throughout the space). In reality, the number of combinations is so high that many sequences should have a similar fitness (Gavrilets, 2010) and the fitness of neighbouring sequences should have a degree of correlation. This should allow for the appearance of low gradient "fitness ridges" connecting fitness peaks (i.e. saddle points) (Fragata et al., 2019; Maynard Smith, 1970). This is difficult to visualise in the classical "peaks and valleys" metaphor of the landscape which might better be thought of as a densely connected network (Fragata et al., 2019).

In my opinion, the representation of a sequence in a Cartesian coordinate system is in itself misleading. I hardly see a biological basis for considering each position in a sequence as a dimension orthogonal to all others. For example, two genes having overlapping functions are hardly orthogonal. Moreover, the axes of this space should not be considered linear. For example, amino acids should not be regularly spaced along a "dimension" but the distance separating them should represent differences in their physio-chemical properties. Therefore, the space between two proteins varying by one residue depends on the type of substitution **(fig. 5.2d)**. The very concept of *genetic interaction* might simply arise from the discrepancy between the expectations of a linear Cartesian coordinate system and the non-linear reality of living cells.

Experimental studies mapping the fitness landscape arising from the combination of mutations tend to converge

towards two observations relevant to higher order interaction screen with CRI-SPA. Firstly, although epistasis exists, the landscape is sufficiently smooth to allow for any sequence to be connected to the global maximum through an upward adaptive walk. That is, a sequence in the landscape would most certainly converge regardless of its starting point (Bank et al., 2016; Chou et al., 2011; Khan et al., 2011). Second, the improvement of a positive mutations is negatively correlated to the fitness of the background in which it is introduced. In simpler words, accumulating beneficial mutations shows a pattern of diminishing returns (Chou et al., 2011; Khan et al., 2011; Kryazhimskiy et al., 2014).

Although this pattern of diminishing return has been reported in studies interested in fitness, it will most certainly also apply to the productivity phenotype. Adding to this difficulty, the operational range of the biosensor might become saturated as top edits are combined in higher order interaction screens. I expect that this saturation in product synthesis and its reporting will challenge the identification of new hits. Indeed, the diminishing signal difference obtained with new hits might become masked by the noise of the screen and the variation within CRI-SPA quadruplicates. As a result, a higher number of false negatives is expected in higher order interaction screens. In this context, intermediate re-arraying of hits to validate their fluorescence and HPLC validation as performed in Chapter 4 will be particularly important.

Besides its fundamental interest, a greedy search of the interaction space must outperform simpler alternatives to justify its value for the metabolic engineering community. For example, combining beneficial edits identified in a first-level interaction screen would be a simpler strategy and the classical metabolic engineering approach (Yang et al., 2018). Outperforming this simple benchmark should be set as a ground goal. In addition, if higher order interaction screens recommend the same hits as the first order screens (i.e. no idiosyncrasy), their use for metabolic engineering would be difficult to advocate.

If diminishing returns curtail the benefit of higher order genetic interactions, the combinatorial space could be opened to factor-in environmental conditions which have been left out of consideration in this entire thesis. A growing body of literature is investigating the effects of environmental conditions on genetic variants (GxE) and on gene:gene interactions (GxGxE) (Bank, 2022). For example, Hall et al. showed that a simple interaction landscape resulting from the combination of 5 beneficial point mutations *E. coli* adopted different topologies in various growth conditions (Hall et al., 2019). Similarly to double mutant interactions, GxE interactions are not randomly distributed. Knock-out of genes with similar functions tend to have related condition-dependent growth profiles and a minority of gene hubs concentrate a large number of environment interactions (Nichols et al., 2011). This hints that the size of a the library screened for GxE(xE) could be reduced to a subset genes, for example one that is broadly representative of the yeast genome (Kuzmin et al., 2018; Costanzo et al., 2021). Alternatively, growth conditions could be optimised in parallel for a selected number of hits identified by a first genetic interaction screen. A number of computational tools including design of experiment (Gilman et al., 2021) and active learning could be leveraged to direct this search (Borkowski et al., 2020; Radivojević et al., 2020).

In conclusion, this thesis proposes to transform genetic interactions into a source of phenotype improvement for metabolic engineering and synthetic biology. Using CRI-SPA, a new high throughput gene editing method, we have systematically searched for interactions between a metabolic pathway and thousands of modified host loci found in the Yeast Knock Out and Over-Expression libraries. We have shown that the search could be performed with a biosensor and was hence applicable to broad range of colorless chemicals. The high productivity of the top hits identified by CRI-SPA confirms the validity of exploiting interactions in a metabolic engineering context.

Bibliography

Bank, C. (2022), 'Epistasis and adaptation on fitness landscapes', *Annual review of ecology, evolution, and systematics* **53**, 457–479.

Bank, C., Matuszewski, S., Hietpas, R. T. and Jensen, J. D. (2016), 'On the (un) predictability of a large intragenic fitness landscape', *Proceedings of the National Academy of Sciences* **113**(49), 14085–14090.

Borkowski, O., Koch, M., Zettor, A., Pandi, A., Batista, A. C., Soudier, P. and Faulon, J.-L. (2020), 'Large scale active-learning-guided exploration for in vitro protein production optimization', *Nature Communications* **11**(1), 1–8.

Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D. and Marx, C. J. (2011), 'Diminishing returns epistasis among beneficial mutations decelerates adaptation', *Science* **332**(6034), 1190–1192.

Costanzo, M., Hou, J., Messier, V., Nelson, J., Rahman, M., VanderSluis, B., Wang, W., Pons, C., Ross, C., Ušaj, M. et al. (2021), 'Environmental robustness of the global yeast genetic interaction network', *Science* **372**(6542), eabf8424.

Ferreira, R., Skrekas, C., Hedin, A., Sanchez, B. J., Siewers, V., Nielsen, J. and David, F. (2019), 'Model-assisted fine-tuning of central carbon metabolism in yeast through dcas9-based regulation', *ACS Synthetic Biology* **8**(11), 2457–2463.

Fragata, I., Blanckaert, A., Louro, M. A. D., Liberles, D. A. and Bank, C. (2019), 'Evolution in the light of fitness landscape theory', *Trends in ecology & evolution* **34**(1), 69–82.

Gavrilets, S. (2010), 'High-dimensional fitness landscapes and speciation', *Evolution: the extended synthesis* pp. 45–79.

Gilman, J., Walls, L., Bandiera, L. and Menolascina, F. (2021), 'Statistical design of experiments for synthetic biology', *ACS synthetic biology* **10**(1), 1–18.

Hall, A. E., Karkare, K., Cooper, V. S., Bank, C., Cooper, T. F. and Moore, F. B.-G. (2019), 'Environment changes epistasis to alter trade-offs along alternative evolutionary paths', *Evolution* **73**(10), 2094–2105.

Jakočiūnas, T., Bonde, I., Herrgård, M., Harrison, S. J., Kristensen, M., Pedersen, L. E., Jensen, M. K. and Keasling, J. D. (2015), 'Multiplex metabolic pathway engineering using crispr/cas9 in saccharomyces cerevisiae', *Metabolic engineering* **28**, 213–222.

Jakociunas, T., Rajkumar, A. S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C. B., Skjødt, M. L., Nielsen, A. T., Borodina, I., Jensen, M. K. et al. (2015), 'Casemblr: Cas9-facilitated multiloci genomic integration of in vivo assembled dna parts in saccharomyces cerevisiae', *ACS synthetic biology* **4**(11), 1226–1234.

Kauffman, S. and Levin, S. (1987), 'Towards a general theory of adaptive walks on rugged landscapes', *Journal of theoretical Biology* **128**(1), 11–45.

Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E. and Cooper, T. F. (2011), 'Negative epistasis between beneficial mutations in an evolving bacterial population', *Science* **332**(6034), 1193–1196.

Klopfenstein, D., Zhang, L., Pedersen, B. S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C. J., Yunes, J. M., Botvinnik, O., Weigel, M. and others (2018), 'GOATOOLS: A Python library for Gene Ontology analyses', *Scientific reports* **8**(1), 1–17. Publisher: Nature Publishing Group.

Kryazhimskiy, S., Rice, D. P., Jerison, E. R. and Desai, M. M. (2014), 'Global epistasis makes adaptation predictable despite sequence-level stochasticity', *Science* **344**(6191), 1519–1522.

Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Usaj, M. M., Van Leeuwen, J. et al. (2018), 'Systematic analysis of complex genetic interactions', *Science* **360**(6386), eaao1729.

Lian, J., HamediRad, M., Hu, S. and Zhao, H. (2017), 'Combinatorial metabolic engineering using an orthogonal tri-functional crispr system', *Nature communications* **8**(1), 1–9.

Liao, C., Ttofali, F., Slotkowski, R. A., Denny, S. R., Cecil, T. D., Leenay, R. T., Keung, A. J. and Beisel, C. L. (2019), 'Modular one-pot assembly of crispr arrays enables library generation and reveals factors influencing crrna biogenesis', *Nature communications* **10**(1), 1–14.

Maynard Smith, J. (1970), 'Natural selection and the concept of a protein space', *Nature* **225**(5232), 563–564.

McCarty, N. S., Graham, A. E., Studená, L. and Ledesma-Amaro, R. (2020), 'Multiplexed crispr technologies for gene editing and transcriptional regulation', *Nature Communications* **11**(1), 1–13.

Nichols, R. J., Sen, S., Choo, Y. J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K. M., Lee, K. J., Wong, A. et al. (2011), 'Phenotypic landscape of a bacterial cell', *Cell* **144**(1), 143–156.

Radivojević, T., Costello, Z., Workman, K. and Garcia Martin, H. (2020), 'A machine learning automated recommendation tool for synthetic biology', *Nature communications* **11**(1), 4879.

Ryan, O. W., Skerker, J. M., Maurer, M. J., Li, X., Tsai, J. C., Poddar, S., Lee, M. E., DeLoache, W., Dueber, J. E., Arkin, A. P. et al. (2014), 'Selection of chromosomal dna libraries using a multiplex crispr system', *Elife* **3**, e03703.

Tak, Y. E., Kleinstiver, B. P., Nuñez, J. K., Hsu, J. Y., Horng, J. E., Gong, J., Weissman, J. S. and Joung, J. K. (2017), 'Inducible and multiplex gene regulation using crispr–cpf1-based transcription factors', *Nature methods* **14**(12), 1163.

Wright, S. et al. (1932), 'The roles of mutation, inbreeding, crossbreeding, and selection in evolution'.

Yang, D., Kim, W. J., Yoo, S. M., Choi, J. H., Ha, S. H., Lee, M. H. and Lee, S. Y. (2018), 'Repurposing type iii polyketide synthase as a malonyl-coa biosensor for metabolic engineering in bacteria', *Proceedings of the National Academy of Sciences* **115**(40), 9835–9844.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S. et al. (2017), 'Multiplex gene editing by crispr–cpf1 using a single crrna array', *Nature biotechnology* **35**(1), 31.

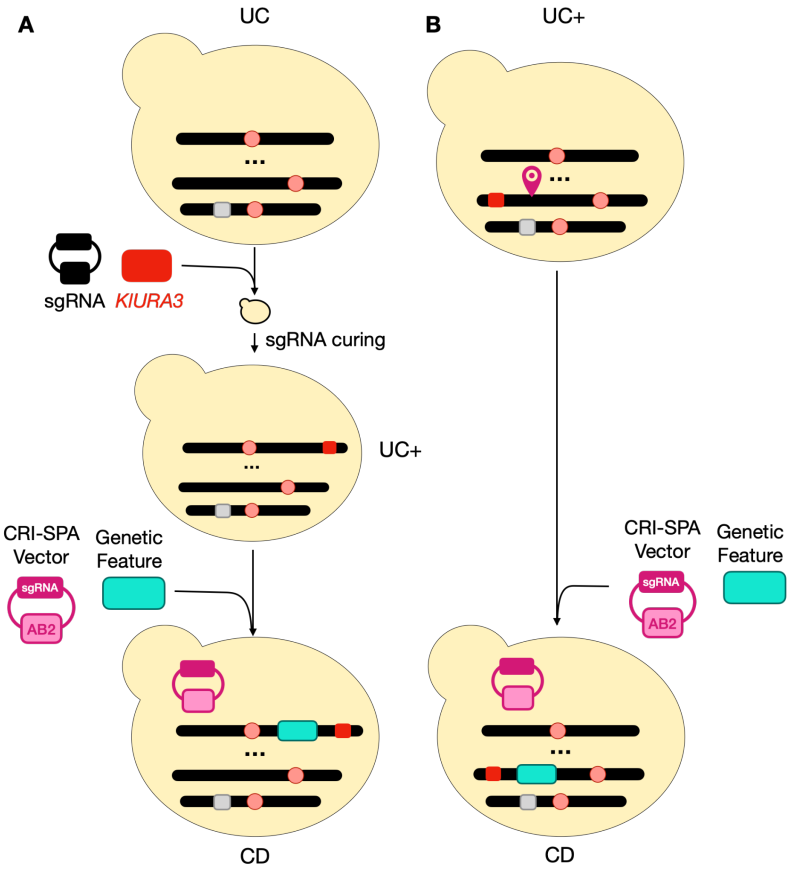
Zhang, Y., Wang, J., Wang, Z., Zhang, Y., Shi, S., Nielsen, J. and Liu, Z. (2019), 'A grna-trna array for crispr-cas9 based rapid multiplexed genome editing in saccharomyces cerevisiae', *Nature communications* **10**(1), 1–10.

Chapter 6

Appendix I: Supplementary Material for: *CRI-SPA: a high-throughput method for systematic genetic editing of yeast libraries*

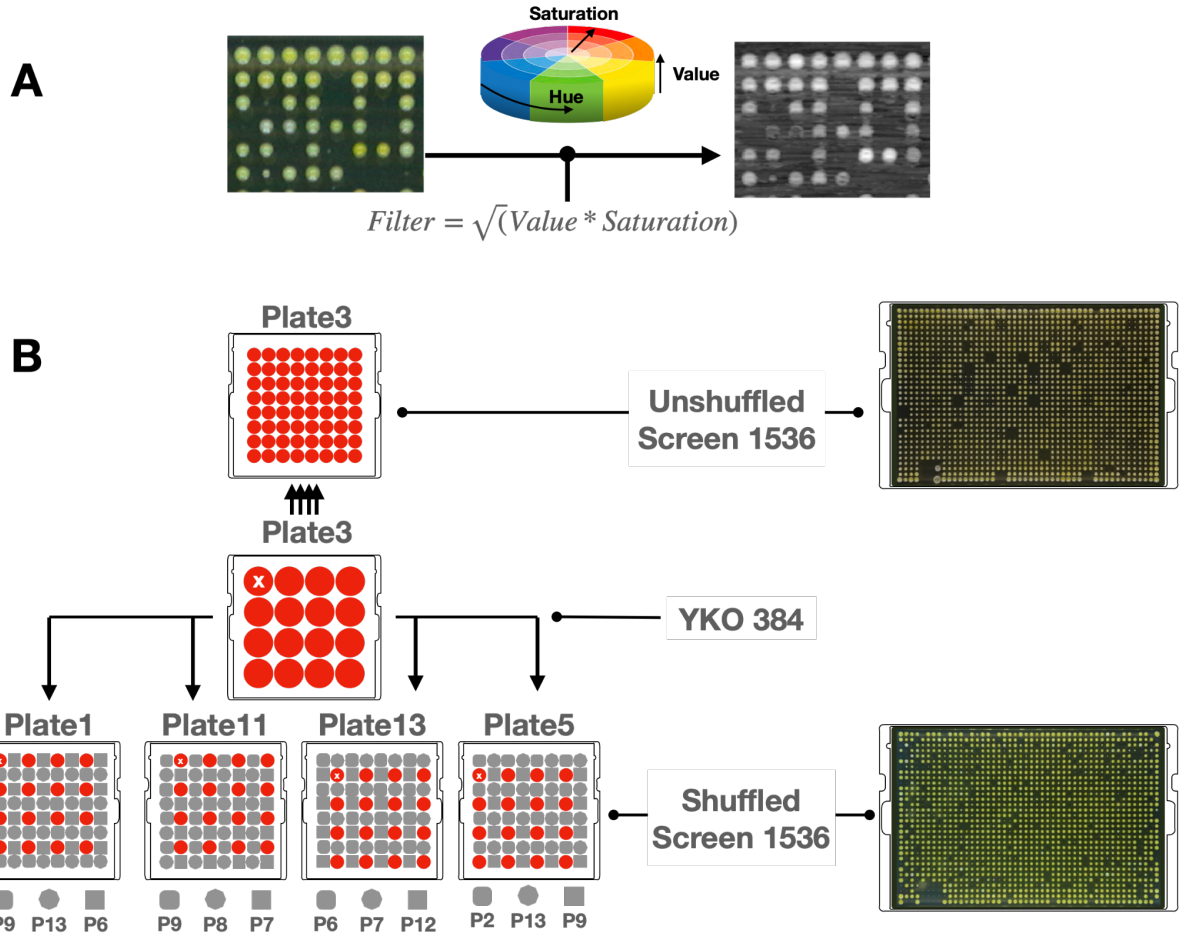
CRI-SPA – a high-throughput method for systematic genetic editing of yeast libraries

Paul Cachera*, Helén Olsson*, Hilde Coumou*, Mads L. Jensen, Benjamín J. Sánchez, Tomas Strucko, Marcel van den Broek, Jean-Marc Daran, Michael K. Jensen, Nikolaus Sonnenschein, Michael Lisby, Uffe H. Mortensen.

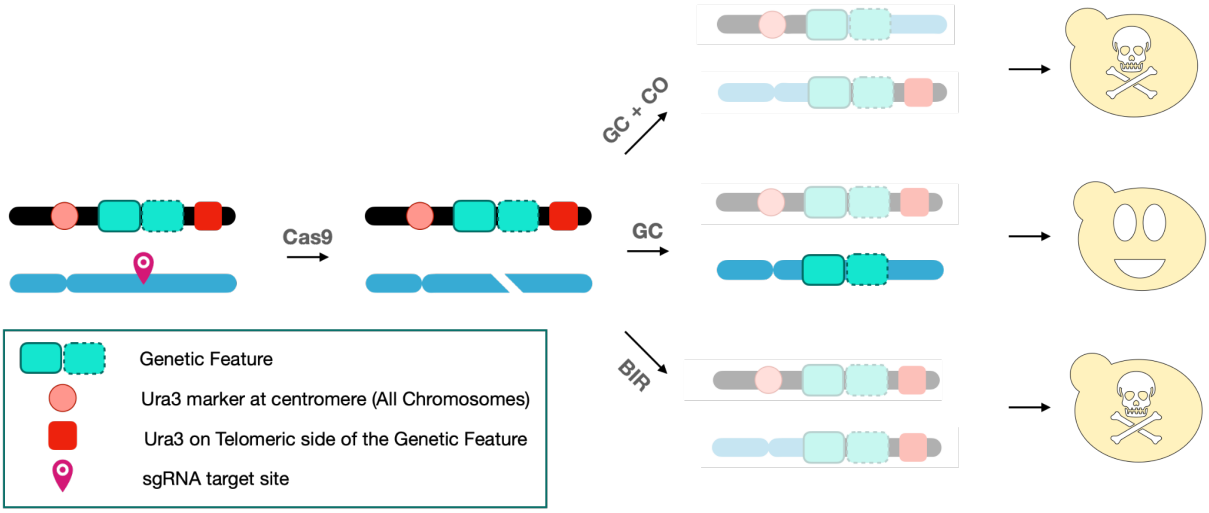


Supplementary Figure S1. CRI-SPA Donor (CD) Strain construction. Construction of a CD strain from a Universal CRI-SPA (UC) depends on the nature of the genetic feature of interest. **(A)** This method employs the basic UC strains and can be used to make any CD strain, but is typically used if the transfer site has not previously been used in a CRI-SPA transfer experiment. In this case, the CD strain needs to be constructed in three steps using a UC strain. Firstly, a *Kl_URA3* marker needs to be incorporated on the telomeric side of the transfer site with the help of an sgRNA

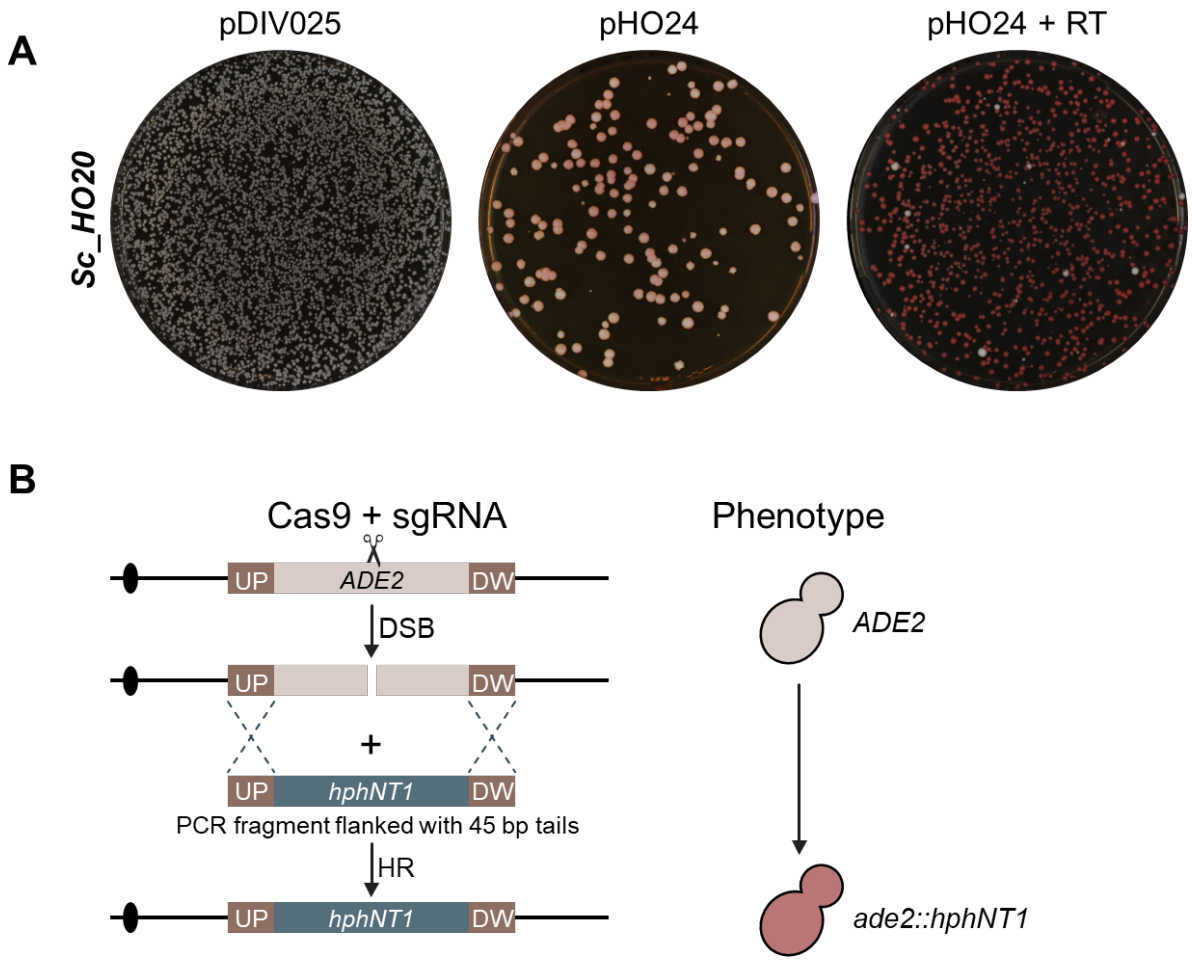
plasmid. Secondly, the strain is cured for the sgRNA plasmid to form an UC+ strain. Thirdly, an appropriate gene-targeting substrate containing the genetic feature is transformed into the UC+ strain together with the relevant CRI-SPA vector to produce the CD strain. **(B)** For construction of a set of CD strains harboring different gene-expression cassettes in the same common insertion site, a common UC+ strain serves as the starting point for construction of all strains in the set to keep strain construction work at a minimum. In this case, construction of the first CD in the set follows track (A) and is constructed in two steps. The UC+ strain created in this process can subsequently be used as a starting point for the construction of the remaining CD strains of the set, which can be made in a single step as shown in panel B. In this manuscript, we describe a UC+ strain, UC+:XII-5, which comes with a pre-integrated *Kl_URA3* marker between the defined expression site XII-5 and the corresponding telomere. Moreover, an efficient CRI-SPA vector for this site is also available (pHO-XII-5 see Supplementary Table S2). For convenience, UC- and UC+:XII-5 strains are available in both mating types (see Supplementary Table S1).



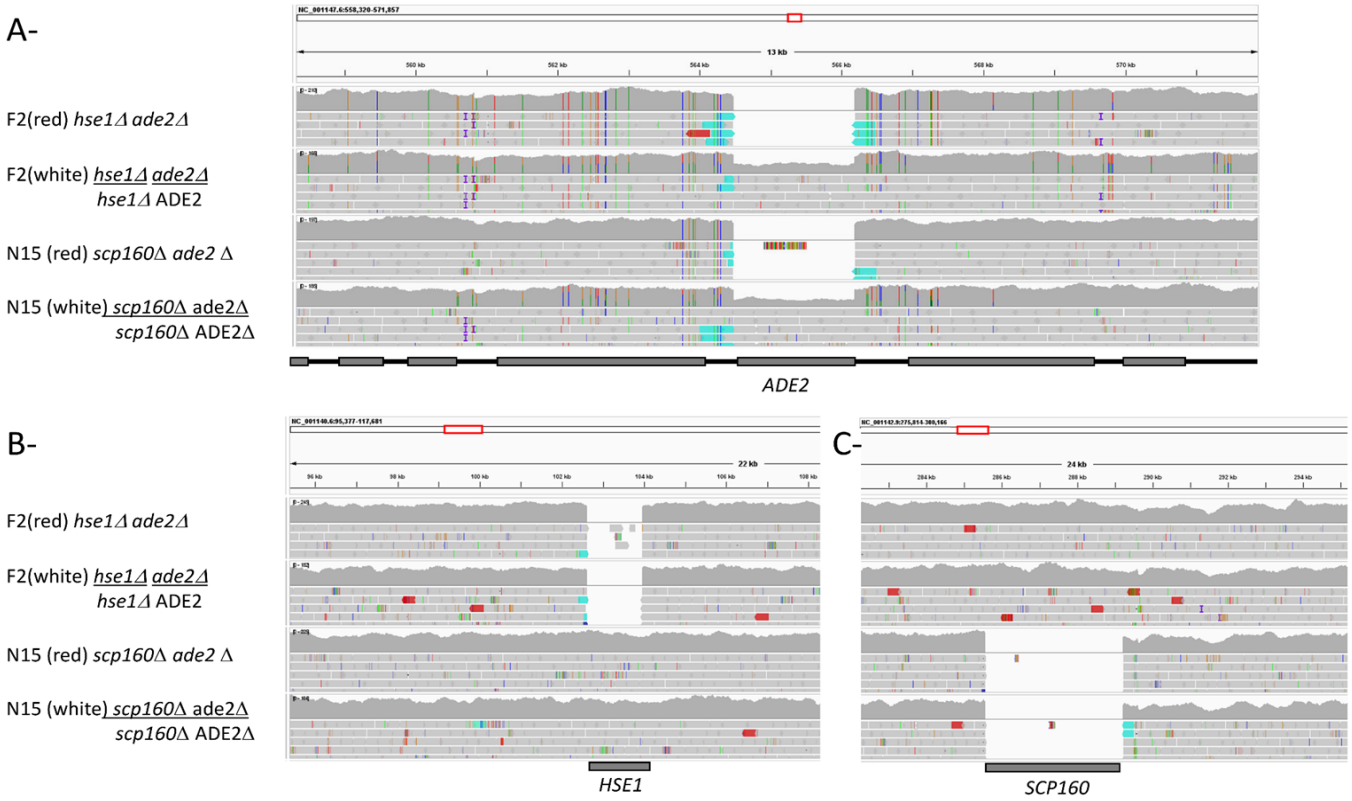
Supplementary Figure S2. Colony Shuffling Scheme and Color FIltering. (A) Left: first version of the screen. During expansion from 386 to 1544 format, individual mutant colonies are pinned as colony quadruplicates in a square of the same plate. As a result, each quadruplicate colony is affected by identical agar quality and similar neighbor effects. Blue square shows YOR1 mutant. Right, second version of the screen. During expansion from 386 to 1544 format, individual mutant colonies are shuffled across random screen-plates. For example, YKO plate 1 is pinned on screen plates 7, 3, 6 and 10. This increases the diversity of neighbors and reduces potential agar gradient biases. **(B)** Heuristic to quantify yellow intensity. Yellowness was quantified as the geometric mean between Value and Saturation in the HSV domain.



Supplementary Figure S3. Cas9 induced crossover and break induced replication produce chimeric strains. In CRI-SPA, transfer of the desirable genetic information from the donor to the target site by homology directed repair of a DNA DSB in the target site is performed by gene conversion (GC). However, homology directed repair of the DNA DSB may also occur via GC accompanied by a crossing over event (GC + CO) or by break induced replication (BIR). In both cases, the final strain after CRI-SPA will contain an undesirable chimeric chromosome consisting of donor and target chromosome sections. In our CRI-SPA system, strains created by GC + CO or BIR are counterselected in step 4 of the CRI-SPA procedure (5-FOA counter selection) as we have incorporated a *K. lactis URA3* marker between the donor site and the telomere; and this *URA3* marker will be transferred to the target chromosome in case that repair of the CRISPR induced DNA DSB involves GC + CO or BIR.

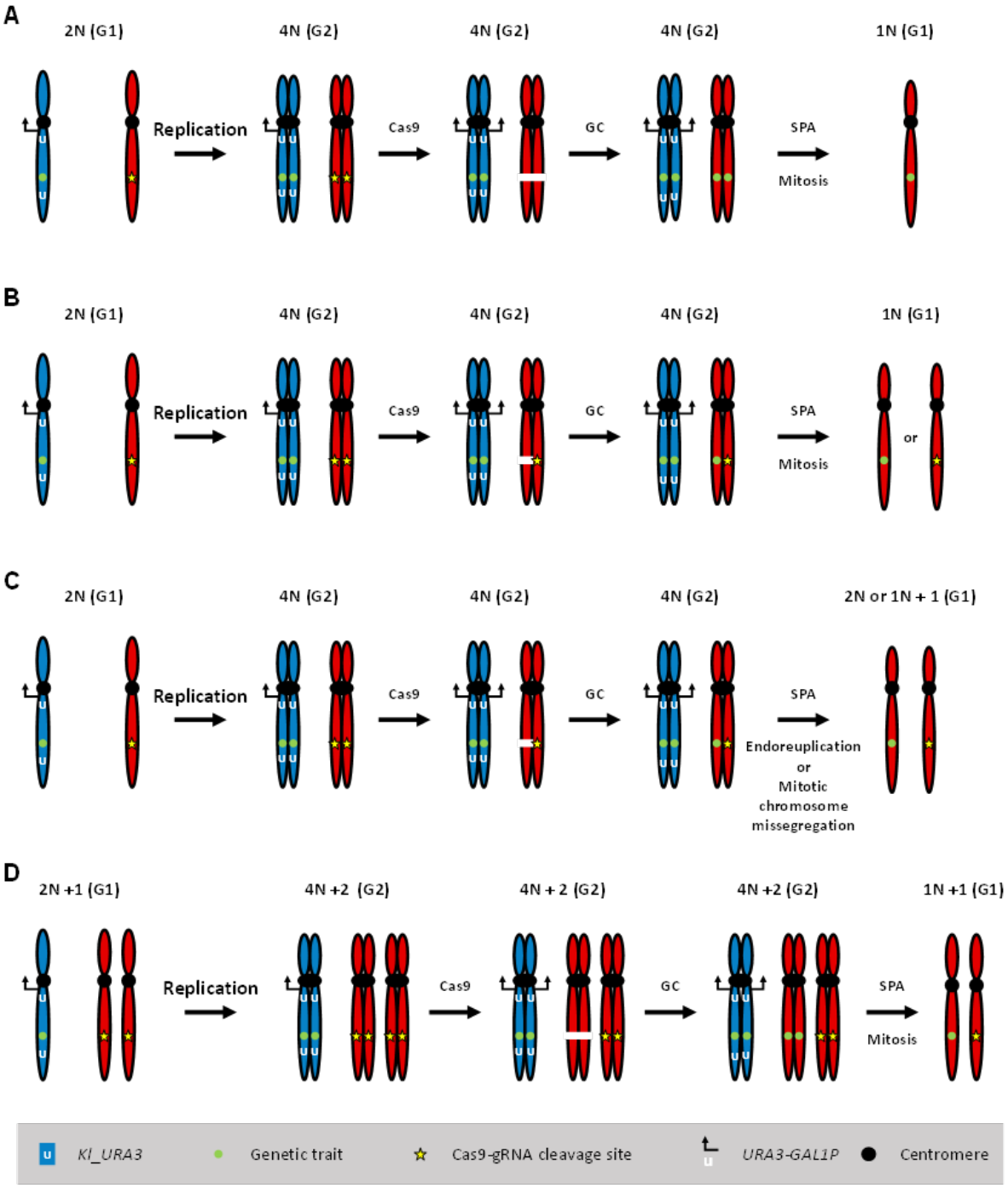


Supplementary Figure S4. TAPE experiment to assess Cas9/gRNA cleavage efficiency of a target protospacer. A TAPE (tool assess protospacer efficiency) experiment (Vanegas et al., 2017) was performed to determine how efficiently the CRISPR nuclease Cas9/*ADE2*-gRNA cuts the *ADE2* locus in a UC strain (UC+XII5α). **(A)** Plates showing the numbers of transformants obtained after transformation of the UC strain with an empty plasmid pDIV025, with pHO24 encoding the *ADE2*-sgRNA, and after co-transformation with pHO24 and a repair template (RT) composed by a *ade2Δ::hphNT1* PCR fragment flanked with 45 bp long UP and DW targeting regions. The amounts of vector DNA used in the three transformation experiments were stoichiometrically the same. (**B**) Cartoon showing HR mediated repair of the Cas9/*ADE2*-gRNA induced DNA DSB in ADE2 using *ade2Δ::hphNT1* as repair template for breaks induced at *ADE2*. The development of a red phenotype resulting from the *ade2Δ::hphNT1* mutation is indicated. The high number of transformants obtained with pDIV025 relative to the number obtained with pHO24 indicates that lethal DNA DSBs are produced by the Cas9/*ADE2-s*gRNA CRISPR nuclease in the UC strain transformed with pHO24. This conclusion was supported by two observations. Firsty, the number of transformants obtained after transformation with pHO24 increases to match the number of transformants obtained with pDIV25 when the repair fragment was included in the reaction allowing for efficient DNA DSB repair by homologous recombination. Secondly, virtually all transformations obtaied in the co-transformation experiment were red as *ADE2* was delete



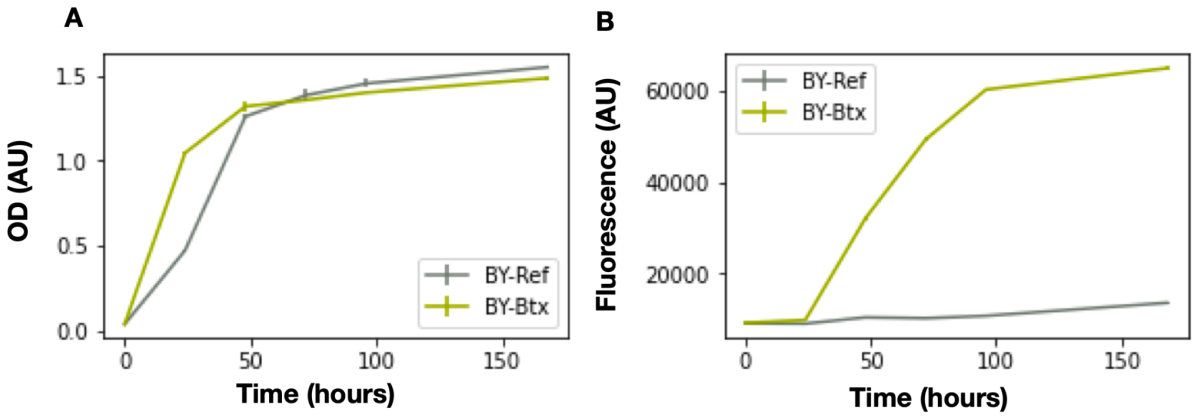
Supplementary Figure S5. Sequencing of selected CRI-SPA strain after CRI-SPA mediated *ade2Δ::hphNT1* cassette transfer from CD-*ade2Δ* strain to plate 9 strains of the YKO library. IGV screenshot of mapping of sequencing reads of red and white isolates of library

clones F2 and N15, which harbor an *HSE1* and *SCP160* deletion, respectively, after CRI-SPA transformation for *ADE2* deletion on genome of the Saccharomyces cerevisiae strain CEN.PK113-7D. **(A)** shows illumina sequencing read mapping at *ADE2* locus. **(B)** shows mapping at *HSE1* locus and **(C)** at locus *SCP160* locus.


A
2N (G1)
Replication
4N (G2)
Cas9
4N (G2)
GC
4N (G2)
SPA
Mitosis
1N (G1)
B
2N (G1)
Replication
4N (G2)
Cas9
4N (G2)
GC
4N (G2)
SPA
Mitosis
1N (G1)
or
C
2N (G1)
Replication
4N (G2)
Cas9
4N (G2)
GC
4N (G2)
SPA
Endoreuplication
or
Mitotic
chromosome
missegregation
2N or 1N + 1 (G1)
D
2N +1 (G1)
Replication
4N +2 (G2)
Cas9
4N + 2 (G2)
GC
4N +2 (G2)
SPA
Mitosis
1N +1 (G1)
KI_URA3
Genetic trait
Cas9-gRNA cleavage site
URA3-GAL1P
Centromere

Supplementary Figure S6. Models explaining background cells generated by CRI-SPA. Background may be produced in cells containing an extra copy of the target recipient chromosome. The extra chromosome may be due to the recipient strain being aneuploid from the beginning, or due to endoreduplication during SPA. If Cas9 only cleaves the target sequence in one of the two target chromosomes during CRI-SPA, CRI-SPA mediated transfer of a genetic trait from a donor chromosome (blue) to the corresponding recipient chromosome (red) is incomplete. Different scenarious are illustrated in the panel below with the diploid strain formed by mating of the CD to a recipient strain during CRI-SPA as a starting point. **(A)** in normal CRI-SPA, Cas9 will cleave both recipient target chromatids in the diploid cells formed by fusion of donor and recipient cells. Repair of the breaks by gene conversion (GC) transfers the genetic trait from a donor chromatid to the corresponding recipient chromatids. Haploidization by SPA generates recipient chromosomes Containing the genetic trait of interest. **(B)** If CRISPR mediated GC takes place after the target locus has been replicated, the possibility exists that only one of the two recipient chromatids may be cleaved by Cas9 in the timeframe of the CRI-SPA process where Cas9 is active. If so, only one of the two recipient chromatids will receive genetic information from the corresponding donor chromatid. In this case cells may contain either a recipient chromosome containing the desired genetic trait, or one that does not. If the genetic trait is selected for, the latter type of cells will die. If the genetic trait is not selected for, the latter type of cells will survive and produce background. **(C)** Same scenario as in B) except that the process does not continue with a perfect mitosis, rather the entire genome or the target chromosome is duplicated as the result of endoreduplication or the number of target chromosomes is doubled due to chromosome missegregation. As a result, cells containing heterozygous recipient disomes are

formed where only one chromosome contains the desired genetic trait. Such cells are expected to survive in the presence or absence of selection for the genetic trait to produce background. **(D)**. The starting recipient cell is aneuploid and contains two target chromosomes of which only one contains the Btx-cassette. Like in the other scenarios, incomplete cleavage by Cas9 produce one modified and one un-modified target chromosome in the final cell.

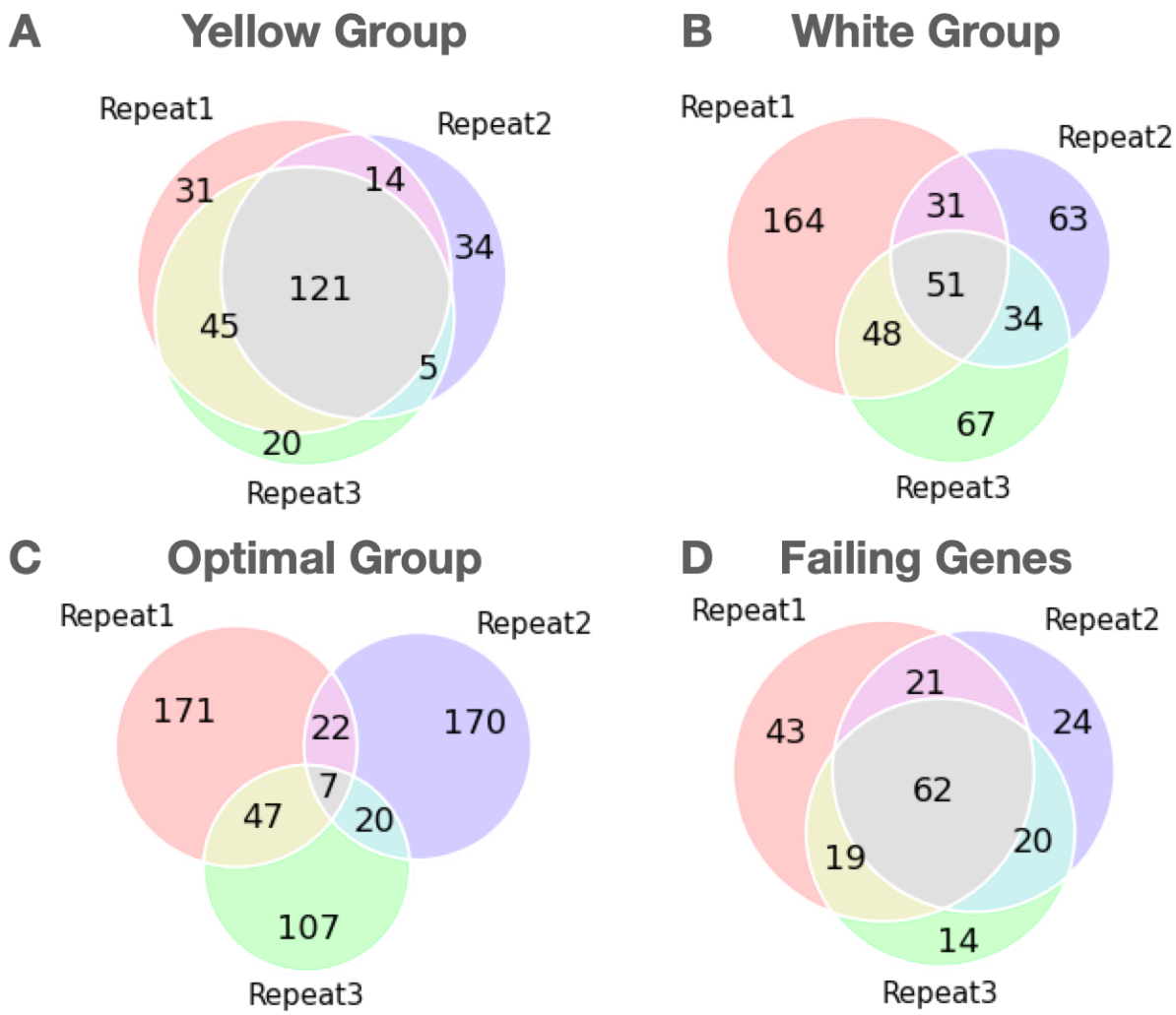


Supplementary Figure S7. Growth profile and Betaxanthin production of BY-Btx.

(A) Growth profile of the betaxanthin producing BY-btx strain and the reference background strain BY-Ref. **(B)** Betaxanthin synthesis for the same samples measured by fluorescence. Error bars are experimental triplicates of the same biological sample.

SEE SUPPLEMENTARY FIGURE S8 DOCUMENT

Supplementary Figure S8. Analysis of CRI-SPA colonies for the presence of unmodified YKO-, CD-, and diploid cells in the absence of selection for the Btx-Cassette during CRI-SPA mediated transfer



Supplementary Figure S9. Screen repeats overlap. Overlap in gene groups between three screen repeats. **(A)** Yellow Group, top betaxanthin (Yellow intensity > 1.2 std above screen mean). **(B)** White Group, lowest betaxanthin (Yellow intensity < 1.2 std below screen mean). **(C)** Cyan group (i.e."cyan group", 0.8 < Yellow intensity < 1.2 std above screen mean). **(D)** Failing genes, genes producing 0 colonies after CRI-SPA procedure.

SUPPLEMENTARY TABLES

SUPPLEMENTARY TABLE S1: Strains Used in this Study

Strain Name	Parent	Genotype	Source
W8164-2B	-	MATα CEN1-16::pGal1-KlURA3 ADE2 can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5	Reid et al., 2011
W8164-2C	-	MATa CEN1-16::pGAL1-KlURA3 ADE2 can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5	Reid et al., 2011
BY4741	-	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Baker Brachmann et al., 1998
UC1α	W8164-2B	MATα CEN1-16::pGAL1-KlURA3 ADE2 can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9(Hs)-tCYC1-loxP-KlLEU2	This study
UC1a	W8164-2C	MATa CEN1-16::pGAL1-KlURA3 ADE2 can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9(Hs)-tCYC1-loxP-KlLEU2	This study
UC+XII5α	W8164-2B	MATα CEN1-16::pGAL1-KlURA3 ADE2 can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-Spcas9(Hs)-tCYC1-loxP-KlLEU2 (dw XII-5 IS) KlURA3	This study
UC+XII5a	W8164-2B	MATa CEN1-16::pGAL1-KlURA3 ADE2 can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9(Hs)-tCYC1-loxP-KlLEU2 (dw XII-5 IS) KlURA3	This study
CD-*ade2Δ*	UCS1α	MATα CEN1-16::pGAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9-tCYC1-loxP-KlLEU2-loxP KlURA3(dw XII-5 IS)	This study
CD-1	**UCS1α**	MATα CEN1-16::pGAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9-tCYC1-loxP-KlLEU2-loxP KlURA3(dw XII-5 IS)	This study
CD-2	UCS1a	MATa CEN1-16::pGAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9-tCYC1-loxP-KlLEU2-loxP KlURA3(dw XII-5 IS)	This study
CD-btx	**UCS2α**	*MAT*α CEN1-16::pGal1-*KlURA3 can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2* (dw XII-5 IS) *KlURA3* XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3*-BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
BY-Btx	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3*-BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX007	BY-Btx	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 **dld2Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3*-BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX008	BY-Btx	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 kgd**2Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3*-BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study

ScBTX009	BY-Btx	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mip1**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX011	BY-Btx	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rcr2**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX012	BY-Btx	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 sgm1**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX013	BY-Btx	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 shy1**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX014	BY-Btx	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 vps34**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX016	BY-Btx	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 yer084W**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX018	BY-Btx	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cox12**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX020	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 **XI-3::KanMX**	This study
ScBTX021	ScBTX020	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 **XI-3::KanMX** XII-5::TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3- BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX022	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 hem25**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX024	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 pro2**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX025	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 vma16**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX027	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 get4**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX028	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 pry1**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX029	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 stv1**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX030	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ycr101C**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX031	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 YLR271W**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX032	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 glr1**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX051	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 qcr10**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX052	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ssm4**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study

ScBTX053	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ubc7**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study
ScBTX055	ScBTX004	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rpn4**Δ0** XII-5::*TADH1-ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3-* BvCYP76AD1-PTPI1-PCCW12-MjDOD-Pgossypii:NatMX:tTEF1gossypii	This study

SUPPLEMENTARY TABLE S2: Plasmids Used in this Study

Plasmid Name	Genetic Element	Yeast Marker	Type
pHO29	pSNR52-gRNA(XII-5)-tSUP4	hphNT1	2μ
pHO-*ADE2*	pSNR52-gRNA(ADE2)-tSUP4, NatMX	hphNT1	2μ
pHO8	p*TEF1-Sp*Cas9-t*CYC1*	*KILEU2*	Integrative, X-3
pHO22	pSNR52-target(ChrXV)-tSUP4-NatMX		Integrative downstream of ADE2
pHO25	pSNR52-target(downstream XII-5)-tSUP4-NatMX		Integrative downstream of XII-5
pBTX1	PSNR52-gRNA.XII-5-TSUP4 ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3- BvCYP76AD1-PTPI1-PCCW12-MjDOD pTEFagossypii::NatMX::tTEFagossypi	NatMX, URA3	Integrative, XII-5
pBTX2	TSUP4ARO7G141S-PTEF1-PPGK1-ARO4K229L-TTDH3- BvCYP76AD1-PTPI1-PCCW12-MjDOD	NatMX, URA3	Integrative, XII-5
pHO25	pSNR52-gRNA-tSUP4	NatMX, URA3	2μ
pHO29	gRNA(XII-5)-hphMX (from pMEL12)	hphMX (from pMEL12)	2μ
pCCM023	HomologyA::Cloning site:: HomologyB	Amp (E.coli)	pUC
pBTX013	Gene::Up KanMX Gene:Dwn Gene=**GLR1**	Amp (E.coli)	pUC
pBTX014	Gene::Up KanMX Gene:Dwn Gene=**HEM25**	Amp (E.coli)	pUC
pBTX016	Gene::Up KanMX Gene:Dwn Gene= **PRO2**	Amp (E.coli)	pUC
pBTX017	Gene::Up KanMX Gene:Dwn Gene= **VMA16**	Amp (E.coli)	pUC
pBTX019	Gene::Up KanMX Gene:Dwn Gene=**GET4**	Amp (E.coli)	pUC
pBTX020	Gene::Up KanMX Gene:Dwn Gene=**PRY1**	Amp (E.coli)	pUC
pBTX021	Gene::Up KanMX Gene:Dwn Gene=**STV1**	Amp (E.coli)	pUC
pBTX022	Gene::Up KanMX Gene:Dwn Gene=**YCR101C**	Amp (E.coli)	pUC
pBTX023	Gene::Up KanMX Gene:Dwn Gene= **YLR271W**	Amp (E.coli)	pUC

SUPPLEMENTARY TABLE S3: Primers Used in this Study

Primers	Sequence	Note
HOP60	GTTTTAGAGCUAGAAATAGCAAG	For exlcusion of protospacer sequence in pMEL12 (or any vector with the same pSNR52/crRNA set-up) while adding USER tails enabling new protospacer to be inserted.
HOP61	GATCATTTAUCTTTCACTGC	For exlcusion of protospacer sequence in pMEL12 (or any vector with the same pSNR52/crRNA set-up) while adding USER tails enabling new protospacer to be inserted.

HOP64	TATAACAATCAAGAAAAACAAGAAAATCGGACAAAACAATCAAGTGCATAGGCCACTAGTGG	For generating a ade2Δ::hygR deletion cassette from PCR with pMEL12
HOP65	ATTTTATAATTATTTGCTGTACAAGTATATCAATAAACTTATATACATCAATAGGCACCTTCG	For generating a ade2Δ::hygR deletion cassette from PCR with pMEL12
HOP89	ACACGCGAUGCTGGAGCTCATAGCTTCA	To amplify pTEF1-Cas9-tCYC1 from pCfB2312 while generating USER tails for cloning into pCfB257/259 linearised by HO_P91 and HO_P92.
HOP90	ACGTGCGAUCTATAGGGCGAATTGGGTAC	To amplify pTEF1-Cas9-tCYC1 from pCfB2312 while generating USER tails for cloning into pCfB257/259 linearised by HO_P91 and HO_P92.
HOP91	ATCGCGTGUGAGCTCGCTGAGGACTT	Opening up pCfB257 or pCfB259. Excluding tADH1 and tCYC1 from the USER cassette, adding USER tails for cloning of new cassette.
HOP92	ATCGCACGUCAGCTGAAGCTTCGTACG	Opening up pCfB257 or pCfB259. Excluding tADH1 and tCYC1 from the USER cassette, adding USER tails for cloning of new cassette.
HOP175	TGATACTGTTCTCATAAACATGTGACTGCATTGGTGGTTGgggttctcgagagctcg	40 bp tail corresponding to UP cas9 target site downstrream ADE2
HOP176	TCCGAGGGTAACTGTGATAGCTTCAAAAGACTTTTAGCATggtctagagatcccaataca	40 bp tail corresponding to DW cas9 target site downstream ADE2
HOP181	GTATACCTTTCATTTACAAAATGTCAAAGACTTTACACAGGGGTTCTCGAGAGCTCG	To amplify KlURA3 from pCfB390 while adding 40 bp tails for integration at genomic site downstream the XII-5 site.
HOP182	CGCAATGACATAGCTTAGAGCAGTCAGAAAAGTGATTGTAGGTCTAGAGATCCCAATACA	To amplify KlURA3 from pCfB390 while adding 40 bp tails for integration at genomic site downstream the XII-5 site.
HOP183	ATTCATGTGATAATTTGACGGTTTTAGAGCT	Cloning gRNA target sequence for downstream XII-5, (+) strand
HOP184	CGTCAAATTATCACATGAATGATCATTTAT	Cloning gRNA target sequence for downstream XII-5, (+) strand
HOP190	AATTGTAGAGACTATCCACAGTTTTAGAGCT	Cloning gRNA target sequence for *ADE2*, (+) strand
HOP191	TGTGGATAGTCTCTACAATTGATCATTTAT	Cloning gRNA target sequence for *ADE2*, (-) strand
HOP217	ACGTGCGAUTCTCGAGAGCTCGTTAAAGC	hphNT1 cassette amplification from pMEL12
HOP218	ACA CGC GAUAGAGATCTGTTTAGCTTGCCTCG	hphNT1 cassette amplification from pMEL12
HOP258	ATTCTGCAGGUAGGGAAAGATATGAG	Construction of plasmid pHO36
HOP259	ACCTGCAGAAUcgagctcgctgaggac	Construction of plasmid pHO36
PR_DIV2288	**CGTGCGAU**TTACTCTTCCAACCTTCTTAGCAAG	Building pBTX2 from pBTX1 excluding the NTC marker
ANT_P494	**CACGCGAU**TTAGGATCCGTCGGTCTTTTGG	Building pBTX2 from pBTX1 excluding the NTC marker
BTX.006	gggacgaggcaagctaaacagatctctagacctaGGATCTTAGTTTCGTTTTTAGTAATG	gene KO: COX12 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.007	agtgtcgaaaacgagctctcgagaacccttaatGTCTATTATACTGTATTTATGTGCTTG	gene KO: COX12 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX

BTX.0103	tcgaaaacgagctctcgagaacccttaatTTTGATGCGGAAACAACGGG	gene KO: QCR10 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.0104	gacgaggcaagctaaacagatctctagacctaCCAAAACATCCAACCGGCT	gene KO: QCR10 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.014	ggcggggacgaggcaagctaaacagatctctagacctaCTTGTTATATGCCACTTGTAGT	gene KO: DLD2 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.015	gtgtcgaaaacgagctctcgagaacccttaatAAATAAAATGCCAAGTTTTATAGTTATC	gene KO: DLD2 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.018	cggcggggacgaggcaagctaaacagatctctagacctaTTTGAAATTTGCAGCCCATAC	gene KO: KGD2 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.019	gtgtcgaaaacgagctctcgagaacccttaatTGTTTAAATAATTATCTGAAGTTGTGTC	gene KO: KGD2 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.022	ggcggggacgaggcaagctaaacagatctctagacctaGCAAAATCTGCATTAGCTTTCG	gene KO: MIP1 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.023	gccatccagtgtcgaaaacgagctctcgagaacccttaatACTTGTCCCTCGTTGCT	gene KO: MIP1 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.030	ggggacgaggcaagctaaacagatctctagacctaTTCAAAACTATACTCTCTGTTTAGC	gene KO: RCR2 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.031	agtgtcgaaaacgagctctcgagaacccttaatTTTTTCGTTTAATGAATTTATTATCGC	gene KO: RCR2 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.034	cgaggcaagctaaacagatctctagacctaCTAGAAATACACTCTTATCTTTAGATTTAG	gene KO: SGM1 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.035	catccagtgtcgaaaacgagctctcgagaacccttaatAGCAAAATTGGTATACCACTAC	gene KO: SGM1 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.038	cggggacgaggcaagctaaacagatctctagacctaTGCTTTCCTCTTTTATTAAATCCT	gene KO: SHY1 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.039	gccatccagtgtcgaaaacgagctctcgagaacccttaatACGGCAAACTCCTCTGC	gene KO: SHY1 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.042	ccggcggggacgaggcaagctaaacagatctctagacctaTGTTGGAGTTAAACCCTTCC	gene KO: VPS34 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.043	ccatccagtgtcgaaaacgagctctcgagaacccttaatTTGGTTGATAATTGGTGCTTC	gene KO: VPS34 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.053	TAGGTCTAGAGATCTGTTTAGC	kanMX
BTX.054	ATTAAGGGTTCTCGAGAGC	kanMX
BTX.057	ccggcggggacgaggcaagctaaacagatctctagacctaGTTTGTAGCATCAGCAACG	gene KO: YER084W Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.058	cagtgtcgaaaacgagctctcgagaacccttaatCTCAGATTTTCAAGAAACAGAATAGC	gene KO: YER084W Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX

BTX.064	ATATATTTTATATATTGCCCTCAAAACT	Verification Primer CAX4
BTX.065	AAACACGCTCTTGAAATTTTTAAAG	Verification Primer CAX4
BTX.066	TCAGGCTGTTGATGAAGGT	Verification Primer COX12
BTX.067	CAGATTTAGGGACTGAGGAA	Verification Primer COX12
BTX.068	TTCTGATTGTAATATAATGACAAATTTTAATC	Verification Primer CUR1
BTX.069	TTCCGTTTTATATGTTTATATTCATTGA	Verification Primer CUR1
BTX.070	CAAGGTCTTCTCTACTGGC	Verification Primer DLD2
BTX.071	ATGAATCAGAAGTTTTAGTTGATAGAA	Verification Primer DLD2
BTX.072	TCAACTATTTAGGATTTTGTAAAGAAAA	Verification Primer KGD2
BTX.073	GTGGGACATATTCGAACGT	Verification Primer KGD2
BTX.074	TACACATCGCCTCCAGA	Verification Primer MIP1
BTX.075	GAATATCCCGTCTAGCGAC	Verification Primer MIP1
BTX.078	GAAGCCTTCACTTTTGCTATC	Verification Primer RCR2
BTX.079	CATTTTGCCTTACTGCATCC	Verification Primer RCR2
BTX.080	GGACAAAAGAATATTCCGATTCA	Verification Primer SGM1
BTX.081	AAAACTGCAGCATCAAATGAC	Verification Primer SGM1
BTX.082	CTCAAAAAATTACTAAGGAAGGC	Verification Primer SHY1
BTX.083	CTCTTATTTTAGGCAACAGAAGC	Verification Primer SHY1
BTX.084	AACTCAAACGTTAATATTATCGGA	Verification Primer VPS34
BTX.085	CCAGCCGAATAAACTTGAC	Verification Primer VPS34
BTX.090	ctgattgagtaaatgcataggc	Verification Primer YER084W
BTX.091	GAGGGTTTTGAACAGGGG	Verification Primer YER084W
BTX.102	CATTTTGGAAACTGCCCCGA	Verification Primer QCR10
BTX.105	CATAGAGATGTGCTGCCCAA	gene KO: QCR10 Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.105	CATAGAGATGTGCTGCCCAA	Verification Primer QCR10
BTX.116	AAGAGGGCuATTTCACTTCGTCCTTAATATTTTAAC	USER primer HEM25_up
BTX.117	CGTGCGAuATGCAAATTTAAATGTAATACGGTA	USER primer HEM25_up
BTX.118	CACGCGAuGTGCACTTATCCATCGACT	USER primer HEM25_dw
BTX.119	ACACAGGCuAAGGTATAAATACGTGTCTTTTTAGT	USER primer HEM25_dw
BTX.120	CGTGCGAuGGCCTATTGGGTATCTGC	USER primer VMA16_up
BTX.121	AAGAGGGCuAACACAGTATACTAGGGGTA	USER primer VMA16_up
BTX.122	ACACAGGCuAGCGCATCCATATTTACTAATAG	USER primer VMA16_dw
BTX.123	CACGCGAuACTGAATAGCTTTTGGGCA	USER primer VMA16_dw
BTX.128	AAGAGGGCuTTGTAAAGTGACCTTGTTTTGAC	USER primer PRO2_up
BTX.129	CGTGCGAuCTAAGGGAAATGGATAGTTTGA	USER primer PRO2_up
BTX.130	CACGCGAuGAGTAGATGTAAGACGTGAAAATG	USER primer PRO2_dw
BTX.131	ACACAGGCuTCCAATTAGCTTTCAGTTTTCC	USER primer PRO2_dw
BTX.132	CGTGCGAuTGGTCAATTCTATGCCTTTTATG	USER primer GLR1_up

BTX.132	CGTGCGAuTGGTCAATTCTATGCCTTTTATG	USER primer GLR1_up
BTX.133	AAGAGGGCuTACGTATAAAACCAACAATTTTATGC	USER primer GLR1_up
BTX.133	AAGAGGGCuTACGTATAAAACCAACAATTTTATGC	USER primer GLR1_up
BTX.134	ACACAGGCuCTTTTACTGCTTTCTAACTTTAAGC	USER primer GLR1_dw
BTX.134	ACACAGGCuCTTTTACTGCTTTCTAACTTTAAGC	USER primer GLR1_dw
BTX.135	CACGCGAuGATTGGTTACTCCTCAAAGATTG	USER primer GLR1_dw
BTX.135	CACGCGAuGATTGGTTACTCCTCAAAGATTG	USER primer GLR1_dw
BTX.149	TATACCTTCCACAAAGGTTTTT	Verification Primer GLR1
BTX.150	TGTTGGTCCAGATTTGGC	Verification Primer GLR1
BTX.151	CCAAATAAGCAAATTTGGTTGG	Verification Primer HEM25
BTX.152	CTCGAATTTAATCTCTCCTCTAATAG	Verification Primer HEM25
BTX.155	CAGTAAAAAGTCCACTTAGCAC	Verification Primer PRO2
BTX.156	GATCTTCTATCAGGAGAGTTAACC	Verification Primer PRO2
BTX.157	GCCGTCTTGACCACAG	Verification Primer VMA16
BTX.158	GATGAGGGAGAGCCAATAA	Verification Primer VMA16
BTX.161	CGTGCGAuTTCTAGTATCTACACTGAAGTCAA	USER primer PRY1_up
BTX.162	AAGAGGGCuTTGTGCTGGAGGGCT	USER primer PRY1_up
BTX.163	ACACAGGCuGAAAAATTTTCGCTCATTTTCGC	USER primer PRY1_dw
BTX.164	CACGCGAuTGCTGTTGTGTACTTAGACG	USER primer PRY1_dw
BTX.165	CGTGCGAuAAAGTATGTCAACCGTCGAG	USER primer GET4_up
BTX.166	AAGAGGGCuTGATGTTTACTTGGTTGTTTTACC	USER primer GET4_up
BTX.167	ACACAGGCuAAAGAGCATATTTGCGAAGGA	USER primer GET4_dw
BTX.168	CACGCGAuAGACATAGAAAGCTATACTCTTATAC	USER primer GET4_dw
BTX.169	CGTGCGAuGGTTGTTAAAATTCTCACATCATCC	USER primer YLR271W_up
BTX.170	AAGAGGGCuTAATTGTTTCATTCTTAGTCTTTAATTTATTT	USER primer YLR271W_up
BTX.171	ACACAGGCuGATGATTTTCTTGAAACTCTCCTTA	USER primer YLR271W_up
BTX.172	CACGCGAuGACAGGTTAAGGTTAAAGGAC	USER primer YLR271W_up
BTX.173	CGTGCGAuTACCTATAATAGCGGCTCG	USER primer YCR101C_up
BTX.174	AAGAGGGCuACCACAGAAAATGGTACAAGA	USER primer YCR101C_up
BTX.175	ACACAGGCuAATGATACTTCGATACTCTGTTTGT	USER primer YCR101C_dw
BTX.176	CACGCGAuTGCCCGTCATCTTGCG	USER primer YCR101C_dw
BTX.177	CGTGCGAuTATTAATAGGTGCTTCTTCCAGG	USER primer STV1_up
BTX.178	AAGAGGGCuCACCTTCGTGGGCC	USER primer STV1_up
BTX.179	ACACAGGCuCACATAGTTTACGGTATATATATGTC	USER primer STV1_dw
BTX.180	CACGCGAuGTGGCATTCCTAGTTAAAATGA	USER primer STV1_dw
BTX.181	AACAATCTCGACATCTTGTAGC	Verification Primer PRY1
BTX.182	GGCGAATCAATGGTAGTTAC	Verification Primer PRY1
BTX.183	CCACTCGATTGAGAACCG	Verification Primer GET4
BTX.184	CCCAACGAGAAACTTGGG	Verification Primer GET4

BTX.185	GACCAGAAGGCTACATGAA	Verification Primer CMG1
BTX.186	TTATCATCAAAATATTGTGAAAAGAGC	Verification Primer CMG1
BTX.187	TATATCCTTACTATCTACTATCCAAC	Verification Primer YCR101C
BTX.188	CATTCAACAGGTGAGAATTTCAA	Verification Primer YCR101C
BTX.189	TTTGCATCTTTTTGTCCACG	Verification Primer STV1
BTX.190	ACTAAGAACCTAAACGGCG	Verification Primer STV1
BTX.200	agagtaccactaattgaatcaaagagactagaagtgtgaaagtcGCAGTCGAGCTCATGC	gene KO: SSM4 repair up Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.201	taaatatgctagcattcattttaaatgtaaggaagaaaacgcctGCATGAGCTCGACTGC	gene KO: SSM4 repair dw Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.202	aaaAGGAACTTCCCTAGTAATAGTGTAATTTGGAAGGGCATAGCGCAGTCGAGCTCATGC	gene KO: UBC7 repair up Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.203	cagttaaaaggaagaccaaatgatcattaacctgctacctgcttGCATGAGCTCGACTGC	gene KO: UBC7 repair dw Upper cap: anneals to gene, ‚Ä®Low. cap: homol. to KanMX
BTX.205	gcatcgattgaggacattg	Verification Primer SSM4
BTX.206	ggatcaaatgcacaacttctc	Verification Primer SSM4

SUPPLEMENTARY TABLE S4: Reversed Engineered Hits

Gene KO	Top/Bottom Hit	Normalised Fitness CRI-SPA score	Normalised Yellowness CRI-SPA score	SGD annotation
WT	NA	0.35	0.22	Betaxanthin making control strain, BY4741. Parent to all KOs
SHY1	Top	-0.82	2.25	Mitochondrial inner membrane protein required for complex IV assembly
DLD2	Top	0.82	0.52	D-2-hydroxyglutarate dehydrogenase, and minor D-lactate dehydrogenase; located in the mitochondrial matrix
MIP1	Top	-1.23	1.51	Mitochondrial DNA polymerase gamma; single subunit of mitochondrial DNA polymerase in yeast
RCR2	Top	0.85	0.94	Vacuolar protein; presumably functions within the endosomal-vacuolar trafficking pathway,
SGM1	Top	0.65	0.65	hypothetical protein; required for wild-type growth rate on galactose and mannose; localizes to COPI coated vesicles and the Golgi apparatus
YER084W	Top	0.0	1.19	hypothetical protein; expressed at both mRNA and protein levels, mitochondrial protein
COX12	Top	-0.75	1.54	Subunit VIb of cytochrome c oxidase; cytochrome c oxidase is also known as respiratory Complex IV and is the terminal member of the mitochondrial inner membrane electron transport chain; required for assembly of cytochrome c oxidase but not required for activity after assembly;
GET4	Top	0.3	0.76	Protein involved in inserting tail-anchored proteins into ER membranes;

PRY1	Top	0.2	1.01	Sterol binding protein involved in the export of acetylated sterols;
STV1	Top	0.8	0.76	Subunit a of the vacuolar-ATPase V0 domain; one of two isoforms (Stv1p and Vph1p); Stv1p is located in V-ATPase complexes of the Golgi and endosomes
YCR101C	Top	0.61	0.59	hypothetical protein; localizes to the membrane fraction;
YLR271W	Top	1.29	0.93	hypothetical protein; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and the nucleus
QCR10	Top	1.63	0.25	Subunit of the ubiquinol-cytochrome c oxidoreductase complex; this complex comprises part of the mitochondrial respiratory chain
SSM4	Top	0.13	1.92	Membrane-embedded ubiquitin-protein ligase and retrotranslocase; ER and inner nuclear membrane localized RING-CH domain E3 ligase involved in ER-associated protein degradation (ERAD);
UBC7	Top	-0.38	2.7	Ubiquitin conjugating enzyme; involved in the ER-associated protein degradation (ERAD) pathway and in the inner nuclear membrane-associated degradation (INMAD) pathway;
RPN4	Top	-0.15	1.15	Transcription factor that stimulates expression of proteasome genes
GLR1	Bottom	0.2	-0.99	Cytosolic and mitochondrial glutathione oxidoreductase; converts oxidized glutathione to reduced glutathione;
HEM25	Bottom	-0.83	-1.65	Mitochondrial glycine transporter; required for the transport of glycine into mitochondria for initiation of heme biosynthesis,
PRO2	Bottom	-1.92	-1.8	Gamma-glutamyl phosphate reductase; catalyzes the second step in proline biosynthesis
VMA16	Bottom	-2.08	-1.57	Subunit c'' of the vacuolar ATPase; v-ATPase functions in acidification of the vacuole
VPS34	Bottom	-1.39	-2.05	Phosphatidylinositol (PI) 3-kinase that synthesizes PI-3-phosphate; forms membrane-associated signal transduction complex with Vps15p to regulate protein sorting;
KGD2	Bottom	-0.23	-1.71	component of the mitochondrial alpha-ketoglutarate dehydrogenase complex, which catalyzes the oxidative decarboxylation of alpha-ketoglutarate to succinyl-CoA in the TCA cycle; phosphorylated

SUPPLEMENTARY TABLE S5: Top Betaxanthin Hits

gene	Yield_Rank	Yield_Score	Size_Score	Description
UBC7	1	3.042	-0.497	Ubiquitin conjugating enzyme; involved in the ER-associated protein degradation (ERAD) pathway and in the inner nuclear membrane-associated degradation (INMAD) pathway; requires Cue1p for recruitment to the ER membrane; proposed to be involved in chromatin assembly
COQ2	2	2.972	-0.675	Para hydroxybenzoate polyprenyl transferase; catalyzes the second step in ubiquinone (coenzyme Q) biosynthesis; human COQ2, mutations in which are implicated in an increased risk of mutiple-system atrophy, can complement a yeast coq2 null mutant
CBP6	3	2.904	-0.48	Mitochondrial protein required for translation of the COB mRNA; forms a complex with Cbp3p that binds to mt ribosomes near the polypeptide tunnel exit and promotes efficient translation of the COB mRNA; Cbp3p-Cbp6p complex also

				interacts with newly synthesized cytochrome b (Cobp) and Cbp4p to promote assembly of Cobp into the cytochrome bc1 complex; Cbp3p-Cbp6p complex is sequestered if assembly of Complex III is blocked, downregulating COB mRNA translation
SHY1	4	2.869	-0.827	Mitochondrial inner membrane protein required for complex IV assembly; associates with complex IV assembly intermediates and complex III/complex IV supercomplexes; similar to human SURF1 involved in Leigh Syndrome; complex IV is also known as cytochrome c oxidase
PET100	5	2.791	-0.886	Chaperone that facilitates the assembly of cytochrome c oxidase; integral to the mitochondrial inner membrane; interacts with a subcomplex of subunits VII, VIIa, and VIII (Cox7p, Cox9p, and Cox8p) but not with the holoenzyme
SEC28	6	2.764	-0.426	Epsilon-COP subunit of the coatomer; regulates retrograde Golgi-to-ER protein traffic; stabilizes Cop1p, the alpha-COP and the coatomer complex; non-essential for cell growth; protein abundance increases in response to DNA replication stress
MRPL7	7	2.717	-0.55	Mitochondrial ribosomal protein of the large subunit; MRPL7 produces both YmL5 and YmL7, which are two different modified forms of the same protein
MRM1	8	2.636	-1.035	Ribose methyltransferase; modifies a functionally critical, conserved nucleotide in mitochondrial 21S rRNA
MTG2	9	2.632	-0.86	Putative GTPase; member of the Obg family; peripheral protein of the mitochondrial inner membrane that associates with the large ribosomal subunit; required for mitochondrial translation, possibly via a role in ribosome assembly
PET54	10	2.629	-0.689	Mitochondrial inner membrane protein; binds to the 5' UTR of the COX3 mRNA to activate its translation together with Pet122p and Pet494p; also binds to the COX1 Group I intron AI5 beta to facilitate exon ligation during splicing
SSM4	11	2.619	0.037	Membrane-embedded ubiquitin-protein ligase; ER and inner nuclear membrane localized RING-CH domain E3 ligase involved in ER-associated protein degradation (ERAD); targets misfolded cytosolic/nucleoplasmic domains of soluble and membrane embedded proteins (ERAD-C) and a transmembrane domain containing substrate (ERAD-M), Sbh2p; C-terminal element (CTE), conserved in human ortholog MARCH10/TEB4, determines substrate selectivity
COQ5	12	2.619	-0.893	2-hexaprenyl-6-methoxy-1,4-benzoquinone methyltransferase; involved in ubiquinone (Coenzyme Q) biosynthesis; localizes to the matrix face of the mitochondrial inner membrane in a large complex with other ubiquinone biosynthetic enzymes; respiratory defect of the null mutant is partially complemented by human COQ5
COQ10	13	2.61	-0.327	Coenzyme Q (ubiquinone) binding protein; functions in the delivery of Q6 to its proper location for electron transport during respiration; START domain protein with homologs in bacteria and eukaryotes; respiratory growth defect of the null mutant is functionally complemented by human COQ10A
CBS2	14	2.61	-1.409	Mitochondrial translational activator of the COB mRNA; interacts with translating ribosomes, acts on the COB mRNA 5'-untranslated leader
PET111	15	2.608	-0.319	Mitochondrial translational activator specific for the COX2 mRNA; located in the mitochondrial inner membrane
ABF2	16	2.603	-0.512	Mitochondrial DNA-binding protein; involved in mitochondrial DNA replication and recombination, member of HMG1 DNA-binding protein family; activity may be regulated by protein kinase A phosphorylation; ABF2 has a paralog, IXR1, that arose from the whole genome duplication; human homolog TFAM can complement yeast abf2 mutant, rescuing the loss-of-mitochondrial DNA phenotype in a yeast abf2 strain
MSS2	17	2.586	-0.681	Peripherally bound inner membrane protein of the mitochondrial matrix; involved in membrane insertion of C-terminus of Cox2p, interacts genetically and physically with Cox18p
PET494	18	2.577	-1.028	Mitochondrial translational activator specific for the COX3 mRNA; acts together with Pet54p and Pet122p; located in the mitochondrial inner membrane
SCO1	19	2.576	-0.884	Copper-binding protein of mitochondrial inner membrane; required for cytochrome c oxidase activity and respiration; may function to deliver copper to cytochrome c oxidase; similar to thioredoxins; SCO1 has a paralog, SCO2, that arose from the whole genome duplication

RRG8	20	2.545	-1.354	hypothetical protein; required for mitochondrial genome maintenance; null mutation results in a decrease in plasma membrane electron transport
MRPL51	21	2.537	-1.363	Mitochondrial ribosomal protein of the large subunit
CBP2	22	2.533	-0.858	Required for splicing of the group I intron bI5 of the COB pre-mRNA; nuclear-encoded mitochondrial protein that binds to the RNA to promote splicing; also involved in but not essential for splicing of the COB bI2 intron and the intron in the 21S rRNA gene
CYC3	23	2.505	-0.656	Cytochrome c heme lyase (holocytochrome c synthase); attaches heme to apo-cytochrome c (Cyc1p or Cyc7p) in mitochondrial intermembrane space; human homolog HCCS implicated in microphthalmia with linear skin defects (MLS), and can complement yeast null mutant
IRC19	24	2.505	-0.827	hypothetical protein; YLL033W is not an essential gene but mutant is defective in spore formation; null mutant displays increased levels of spontaneous Rad52p foci
DIA4	25	2.504	-0.491	Probable mitochondrial seryl-tRNA synthetase; mutant displays increased invasive and pseudohyphal growth
MNE1	26	2.484	-0.522	Protein involved in splicing Group I al5-beta intron from COX1 mRNA; mitochondrial matrix protein
MRH4	27	2.466	-0.769	Mitochondrial ATP-dependent RNA helicase of the DEAD-box family; required for assembly of the large subunit of mitochondrial ribosomes; binds to the large subunit rRNA, 21S_rRNA; localizes to the matrix face of the mitochondrial inner membrane and associates with the large subunit precursor and with mature ribosomes
IFM1	28	2.453	-0.497	Mitochondrial translation initiation factor 2
QCR7	29	2.45	-0.566	Subunit 7 of ubiquinol cytochrome-c reductase (Complex III); Complex III is a component of the mitochondrial inner membrane electron transport chain; oriented facing the mitochondrial matrix; N-terminus appears to play a role in complex assembly
MRPL8	30	2.449	-0.629	Mitochondrial ribosomal protein of the large subunit
MLS1	31	2.437	-0.677	Malate synthase, enzyme of the glyoxylate cycle; involved in utilization of non-fermentable carbon sources; expression is subject to carbon catabolite repression; localizes in peroxisomes during growth on oleic acid, otherwise cytosolic; can accept butyryl-CoA as acyl-CoA donor in addition to traditional substrate acetyl-CoA
CUE3	32	2.429	-0.428	hypothetical protein; has a CUE domain that binds ubiquitin, which may facilitate intramolecular monoubiquitination
SLS1	33	2.423	-0.696	Mitochondrial membrane protein; coordinates expression of mitochondrially-encoded genes; may facilitate delivery of mRNA to membrane-bound translation machinery
PET130	34	2.404	-0.807	Protein required for respiratory growth; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies
COX18	35	2.387	-0.9	Protein required for membrane insertion of C-terminus of Cox2p; mitochondrial integral inner membrane protein; interacts genetically and physically with Mss2p and Pnt1p; similar to S. cerevisiae Oxa1, N. crassa Oxa2p, and E. coli YidC; respiratory defect of the null mutant is functionally complemented by human COX18 carrying the N-terminal 54 amino acids of S. cerevisiae Cox18p
COX23	36	2.386	-0.649	Protein that functions in mitochondrial copper homeostasis; mitochondrial intermembrane space protein; essential for functional cytochrome oxidase expression; homologous to Cox17p; contains twin cysteine-x9-cysteine motifs
HAP2	37	2.383	-0.38	Subunit of the Hap2p/3p/4p/5p CCAAT-binding complex; complex is heme-activated and glucose-repressed; complex is a transcriptional activator and global regulator of respiratory gene expression; contains sequences sufficient for both complex assembly and DNA binding; respiratory defect of the null mutant is functionally complemented by human NFYA
COQ3	38	2.382	-0.695	O-methyltransferase; catalyzes two different O-methylation steps in ubiquinone (Coenzyme Q) biosynthesis; component of a mitochondrial ubiquinone-synthesizing complex; phosphoprotein

COQ4	39	2.379	-0.98	Protein with a role in ubiquinone (Coenzyme Q) biosynthesis; possibly functioning in stabilization of Coq7p; located on matrix face of mitochondrial inner membrane; component of a mitochondrial ubiquinone-synthesizing complex; human homolog COQ4 can complement yeast coq4 null mutant
PET112	40	2.375	-1.213	Subunit of the trimeric GatFAB AmidoTransferase(AdT) complex; involved in the formation of Q-tRNAQ; mutation is functionally complemented by the bacterial GatB ortholog
MSR1	41	2.362	-1.131	Mitochondrial arginyl-tRNA synthetase; mutations in human ortholog are associated with pontocerebellar hypoplasia type 6; MSR1 has a paralog, YDR341C, that arose from the whole genome duplication
COQ6	42	2.359	-0.619	Flavin-dependent monooxygenase involved in ubiquinone biosynthesis; responsible for hydroxylation at position C5 and deamination at C4 during ubiquinone (Coenzyme Q) biosynthesis; localizes to matrix face of mitochondrial inner membrane in a large complex with other ubiquinone biosynthetic enzymes; human homolog COQ6 can complement yeast null mutant and is implicated in steroid-resistant nephrotic syndrome (SRNS)
CCM1	43	2.357	-0.696	Mitochondrial 15S rRNA-binding protein; required for intron removal of COB and COX1 pre-mRNAs; has separable roles in stabilizing mitochondrial 15S rRNA and in maturation of the COB and COX1 mRNAs; contains pentatricopeptide repeat (PPR) motifs; mutant is respiratory deficient and has defective plasma membrane electron transport
COX19	44	2.353	-0.773	Protein required for cytochrome c oxidase assembly; located in the cytosol and mitochondrial intermembrane space; putative copper metallochaperone that delivers copper to cytochrome c oxidase; contains twin cysteine-x9-cysteine motifs
CUE1	45	2.348	-0.202	Ubiquitin-binding protein; ER membrane protein that recruits and integrates the ubiquitin-conjugating enzyme Ubc7p into ER membrane-bound ubiquitin ligase complexes that function in the ER-associated degradation (ERAD) pathway for misfolded proteins; contains a CUE domain that binds ubiquitin to facilitate intramolecular monoubiquitination and to promote diubiquitin elongation, facilitating polyubiquitin chain formation
MRPL32	46	2.343	-1.177	Mitochondrial ribosomal protein of the large subunit; protein abundance increases in response to DNA replication stress
CYT2	47	2.337	-0.941	Cytochrome c1 heme lyase; involved in maturation of cytochrome c1, which is a subunit of the mitochondrial ubiquinol-cytochrome-c reductase; links heme covalently to apocytochrome c1; human homolog HCCS can complement yeast cyt2 null mutant
MRPL38	48	2.335	-1.015	Mitochondrial ribosomal protein of the large subunit; appears as two protein spots (YmL34 and YmL38) on two-dimensional SDS gels; protein abundance increases in response to DNA replication stress
MSF1	49	2.326	-0.985	Mitochondrial phenylalanyl-tRNA synthetase; active as a monomer, unlike the cytoplasmic subunit which is active as a dimer complexed to a beta subunit dimer; similar to the alpha subunit of E. coli phenylalanyl-tRNA synthetase
EXO5	50	2.326	-0.697	Mitochondrial 5'-3' exonuclease and sliding exonuclease; required for mitochondrial genome maintenance; distantly related to the RecB nuclease domain of bacterial RecBCD recombinases; may be regulated by the transcription factor Ace2
MEF2	51	2.321	-0.801	Mitochondrial elongation factor involved in translational elongation
CBT1	52	2.314	-0.529	Protein involved in 5' RNA end processing; substrates include mitochondrial COB, 15S_rRNA, and RPM1 transcripts; may also have a role in 3' end processing of the COB pre-mRNA; displays genetic interaction with cell cycle-regulated kinase Dbf2p
YER140W	53	2.311	-1.029	Integral membrane protein of the ER; forms an ER-membrane associated protein complex with Slp1p; identified along with SLP1 in a screen for mutants defective in the unfolded protein response (UPR); proposed to function in the folding of integral membrane proteins; interacts genetically with MPS3; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies

CAT5	54	2.302	-1.198	Protein required for ubiquinone (Coenzyme Q) biosynthesis; localizes to the matrix face of the mitochondrial inner membrane in a large complex with ubiquinone biosynthetic enzymes; required for gluconeogenic gene activation
MIP1	55	2.301	-1.267	Mitochondrial DNA polymerase gamma; single subunit of mitochondrial DNA polymerase in yeast, in contrast to metazoan complex of catalytic and accessory subunits; polymorphic in yeast, petites occur more frequently in some lab strains; human ortholog POLG complements yeast mip1 mutant; mutations in human POLG associated with Alpers-Huttenlocher syndrome (AHS), progressive external ophthalmoplegia (PEO), parkinsonism, other mitochondrial diseases
MRPL17	56	2.295	-1.013	Mitochondrial ribosomal protein of the large subunit
TDH1	57	2.283	-0.565	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), isozyme 1; involved in glycolysis and gluconeogenesis; tetramer that catalyzes the reaction of glyceraldehyde-3-phosphate to 1,3 bis-phosphoglycerate; detected in the cytoplasm and cell wall; protein abundance increases in response to DNA replication stress; GAPDH-derived antimicrobial peptides secreted by S. cerevisiae are active against a wide variety of wine-related yeasts and bateria
SLM3	58	2.283	-0.83	tRNA-specific 2-thiouridylase; responsible for 2-thiolation of the wobble base of mitochondrial tRNAs; human homolog TRMU is implicated in myoclonus epilepsy associated with ragged red fibers (MERRF), and can complement yeast null mutant
SHM1	60	2.268	-1.105	Mitochondrial serine hydroxymethyltransferase; converts serine to glycine plus 5,10 methylenetetrahydrofolate; involved in generating precursors for purine, pyrimidine, amino acid, and lipid biosynthesis; reverse reaction generates serine
MRP20	59	2.268	-1.52	Mitochondrial ribosomal protein of the large subunit
RRG9	61	2.266	-1.321	hypothetical protein; null mutant lacks mitochondrial DNA and cannot grow on glycerol; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies
MRPS5	62	2.261	-1.451	Mitochondrial ribosomal protein of the small subunit
COQ9	63	2.252	-1.131	Protein required for ubiquinone biosynthesis and respiratory growth; localizes to matrix face of mitochondrial inner membrane in a large complex with ubiquinone biosynthetic enzymes; ubiquinone is also known as coenzyme Q; human homolog COQ9 can complement yeast coq9 null mutant
AMS1	64	2.244	-0.932	Vacuolar alpha mannosidase; involved in free oligosaccharide (fOS) degradation; delivered to the vacuole in a novel pathway separate from the secretory pathway
PIM1	65	2.226	-1.184	ATP-dependent Lon protease; involved in degradation of misfolded proteins in mitochondria; required for biogenesis and maintenance of mitochondria
MRPL9	66	2.225	-1.554	Mitochondrial ribosomal protein of the large subunit
COR1	67	2.214	-1.059	Core subunit of the ubiquinol-cytochrome c reductase complex; the ubiquinol-cytochrome c reductase complex (bc1 complex) is a component of the mitochondrial inner membrane electron transport chain
CBS1	68	2.212	-0.634	Mitochondrial translational activator of the COB mRNA; membrane protein that interacts with translating ribosomes, acts on the COB mRNA 5'-untranslated leader
MRPL11	70	2.207	-1.275	Mitochondrial ribosomal protein of the large subunit; localizes to vacuole in response to H2O2
HER2	69	2.207	-1.548	Subunit of the trimeric GatFAB AmidoTransferase(AdT) complex; involved in the formation of Q-tRNAQ; required for remodeling of ER caused by Hmg2p overexpression; similar to bacterial GatA glutamyl-tRNA amidotransferase
MRPL23	71	2.187	-1.071	Mitochondrial ribosomal protein of the large subunit; localizes to vacuole in response to H2O2
QCR2	72	2.176	-0.7	Subunit 2 of ubiquinol cytochrome-c reductase (Complex III); Complex III is a component of the mitochondrial inner membrane electron transport chain; phosphorylated; transcription is regulated by Hap1p, Hap2p/Hap3p, and heme
MRP51	73	2.175	-1.058	Mitochondrial ribosomal protein of the small subunit; MRP51 exhibits genetic interactions with mutations in the COX2 and COX3 mRNA 5'-untranslated leader sequences

MTF2	74	2.174	-0.928	Mitochondrial protein that interacts with mitochondrial RNA polymerase; interacts with an N-terminal region of mitochondrial RNA polymerase (Rpo41p) and couples RNA processing and translation to transcription
QRI5	76	2.166	-0.872	Mitochondrial inner membrane protein; required for accumulation of spliced COX1 mRNA; may have an additional role in translation of COX1 mRNA
COX6	75	2.166	-0.797	Subunit VI of cytochrome c oxidase (Complex IV); Complex IV is the terminal member of the mitochondrial inner membrane electron transport chain; expression is regulated by oxygen levels
PET117	77	2.164	-0.752	Protein required for assembly of cytochrome c oxidase
COX7	78	2.149	-0.525	Subunit VII of cytochrome c oxidase (Complex IV); Complex IV is the terminal member of the mitochondrial inner membrane electron transport chain
RSM24	79	2.138	-1.767	Mitochondrial ribosomal protein of the small subunit
MRPL3	80	2.131	-0.456	Mitochondrial ribosomal protein of the large subunit; located in close proximity to the polypeptide exit channel of the ribosome; mutations in human homolog MRPL44 cause childhood cardiomyopathy; human MRPL44 deficiency results in inefficient assembly of the mitochondrial ribosome, and in tissue-specific respiratory chain deficiency, manifesting as either Complex I+Complex IV or Complex IV deficiency, depending on a cell type
AIM10	81	2.117	-1.495	Protein with similarity to tRNA synthetases; non-tagged protein is detected in purified mitochondria; null mutant is viable and displays elevated frequency of mitochondrial genome loss
MSM1	82	2.107	-0.774	Mitochondrial methionyl-tRNA synthetase (MetRS); functions as a monomer in mitochondrial protein synthesis; functions similarly to cytoplasmic MetRS although the cytoplasmic form contains a zinc-binding domain not found in Msm1p
PET123	83	2.093	-1.199	Mitochondrial ribosomal protein of the small subunit; PET123 exhibits genetic interactions with PET122, which encodes a COX3 mRNA-specific translational activator
MSD1	84	2.087	-0.755	Mitochondrial aspartyl-tRNA synthetase; required for acylation of aspartyl-tRNA; yeast and bacterial aspartyl-, asparaginyl-, and lysyl-tRNA synthetases contain regions with high sequence similarity, suggesting a common ancestral gene
CBP4	85	2.07	0.02	Mitochondrial protein required for assembly of cytochrome bc1 complex; interacts with the Cbp3p-Cbp6p complex and newly synthesized cytochrome b (Cobp) to promote assembly of Cobp into the cytochrome bc1 complex
CYT1	86	2.068	-1.16	Cytochrome c1; component of the mitochondrial respiratory chain; expression is regulated by the heme-activated, glucose-repressed Hap2p/3p/4p/5p CCAAT-binding complex
MRPL13	87	2.066	-1.258	Mitochondrial ribosomal protein of the large subunit; not essential for mitochondrial translation
MSW1	88	2.063	-1.232	Mitochondrial tryptophanyl-tRNA synthetase
HAP5	89	2.054	-0.327	Subunit of the Hap2p/3p/4p/5p CCAAT-binding complex; complex is heme-activated and glucose repressed; complex is a transcriptional activator and global regulator of respiratory gene expression; required for assembly and DNA binding activity of the complex
ATP11	90	2.052	-0.832	Molecular chaperone; required for the assembly of alpha and beta subunits into the F1 sector of mitochondrial F1F0 ATP synthase; N-terminally propionylated in vivo
OCT1	91	2.045	-1.439	Mitochondrial intermediate peptidase; cleaves destabilizing N-terminal residues of a subset of proteins upon import, after their cleavage by mitochondrial processing peptidase (Mas1p-Mas2p); may contribute to mitochondrial iron homeostasis
QCR8	92	2.041	-1.648	Subunit 8 of ubiquinol cytochrome-c reductase (Complex III); Complex III is a component of the mitochondrial inner membrane electron transport chain; oriented facing the intermembrane space; expression is regulated by Abf1p and Cpf1p
MRPS35	93	2.037	-1.175	Mitochondrial ribosomal protein of the small subunit; null mutant does not grow on glycerol, is sensitive to 2,4-dichlorophenol, and accumulates large lipid droplets

COX10	94	2.036	-0.865	Heme A:farnesyltransferase; catalyzes first step in conversion of protoheme to heme A prosthetic group required for cytochrome c oxidase activity; human ortholog COX10 can complement yeast cox10 null mutant; human ortholog COX10 is associated with mitochondrial disorders
SEM1	95	2.026	-1.384	19S proteasome regulatory particle lid subcomplex component; role in Ub-dependent proteolysis and proteasome stability; involved in TREX-2 mediated mRNA export, and in the prevention of transcription-associated genome instability; ubiquitinated by Nedd4-like E3-ligase, Rsp5p; human ortholog DSS1, a BRCA1 binding protein implicated in cancer, complements the yeast null; drives trinucleotide repeat expansion; protein abundance increases in response to DNA replication stress
TUF1	96	2.025	-1.595	Mitochondrial translation elongation factor Tu (EF-Tu); involved in fundamental pathway of mtDNA homeostasis; comprises both GTPase and guanine nucleotide exchange factor activities, while these activities are found in separate proteins in S. pombe and humans; rare mutations in human mitochondrial elongation factor Tu (EFTu) associated with severe lactic acidosis, rapidly progressive fatal encephalopathy, severe infantile macrocystic leukodystrophy with micropolygyria
MSS51	97	2.017	-0.727	Specific translational activator for the mitochondrial COX1 mRNA; loosely associated with the matrix face of the mitochondrial inner membrane; localizes to vacuole membrane in response to H2O2; influences both COX1 mRNA translation and Cox1p assembly into cytochrome c oxidase; binds to heme B, which may be a mechanism for sensing oxygen levels in order to regulate cytochrome c oxidase biogenesis
COX12	98	2.015	-0.505	Subunit VIb of cytochrome c oxidase; cytochrome c oxidase is also known as respiratory Complex IV and is the terminal member of the mitochondrial inner membrane electron transport chain; required for assembly of cytochrome c oxidase but not required for activity after assembly; phosphorylated; easily released from the intermembrane space, suggesting a loose association with Complex IV
IMG2	99	2.009	-1.157	Mitochondrial ribosomal protein of the large subunit; conserved in metazoa, with similarity to human mitochondrial ribosomal protein MRPL49
MRPL16	100	2.004	-1.273	Mitochondrial ribosomal protein of the large subunit; homologous to bacterial L16 ribosomal protein; synthetic lethality with hac1 mutation suggests a possible role in synthesis of precursors for protein glycosylation
MRPL20	102	1.993	-1.31	Mitochondrial ribosomal protein of the large subunit
BCS1	101	1.993	-0.592	Protein translocase and chaperone required for Complex III assembly; member of the AAA ATPase family; forms a homo-oligomeric complex in the mitochondrial inner membrane that translocates the C-terminal domain of Rip1p from the matrix across the inner membrane and delivers it to an assembly intermediate of respiratory Complex III; also required for assembly of the Qcr10p subunit; mutation is functionally complemented by human homolog BCS1L, linked to neonatal diseases
POR1	103	1.99	-1.249	Mitochondrial porin (voltage-dependent anion channel); outer membrane protein required for maintenance of mitochondrial osmotic stability and mitochondrial membrane permeability; couples the glutathione pools of the intermembrane space (IMS) and the cytosol; interacts with Om45 and Om14 in the outer membrane; phosphorylated; protein abundance increases in response to DNA replication stress
MRP49	104	1.989	-0.538	Mitochondrial ribosomal protein of the large subunit; not essential for mitochondrial translation
YGR102C	105	1.985	-1.54	Subunit of the trimeric GatFAB AmidoTransferase(AdT) complex; involved in the formation of Q-tRNAQ; transposon insertion mutant is salt sensitive and null mutant has growth defects; non-tagged protein is detected in purified mitochondria
HAP3	106	1.967	-0.26	Subunit of the Hap2p/3p/4p/5p CCAAT-binding complex; complex is heme-activated and glucose-repressed; complex is a transcriptional activator and global regulator of respiratory gene expression; contains sequences contributing to both complex assembly and DNA binding
MRPL10	107	1.955	-1.466	Mitochondrial ribosomal protein of the large subunit; appears as two protein spots (YmL10 and YmL18) on two-dimensional SDS gels

MRPL37	108	1.939	-1.716	Mitochondrial ribosomal protein of the large subunit
QCR9	109	1.92	-0.362	Subunit 9 of ubiquinol cytochrome-c reductase (Complex III); Complex III is a component of the mitochondrial inner membrane electron transport chain; required for electron transfer at the ubiquinol oxidase site of the complex
MHR1	110	1.911	-1.293	Mitochondrial ribosomal protein of the large subunit; also involved in homologous recombination in mitochondria; required for recombination-dependent mtDNA partitioning; involved in stimulation of mitochondrial DNA replication in response to oxidative stress
MRPL1	111	1.909	-0.802	Mitochondrial ribosomal protein of the large subunit
PET309	112	1.906	-1.467	Specific translational activator for the COX1 mRNA; binds to the COX1 mRNA; also influences stability of intron-containing COX1 primary transcripts; localizes to the mitochondrial inner membrane; contains 12 pentatricopeptide repeats (PPRs)
SLM5	113	1.898	-1.4	Mitochondrial asparaginyl-tRNA synthetase
MEF1	114	1.892	-1.014	Mitochondrial elongation factor involved in translational elongation
MRPL49	115	1.877	-1.347	Mitochondrial ribosomal protein of the large subunit
SWS2	116	1.866	-1.373	Putative mitochondrial ribosomal protein of the small subunit; has similarity to E. coli S13 ribosomal protein; participates in controlling sporulation efficiency; localizes to vacuole in response to H2O2
QRI7	117	1.859	-0.578	Protein involved in threonylcarbamoyl adenosine biosynthesis; Sua5p and Qri7p are necessary and sufficient for RNA t6A modification in vitro; highly conserved mitochondrial protein; essential for t6A modification of mitochondrial tRNAs that decode ANN codons; similar to Kae1p and E. coli YgjD, both of which are also required for tRNA t6A modification; when directed to the cytoplasm, complements the essential function of Kae1p in the KEOPS complex
ATP22	118	1.849	-0.917	Specific translational activator for the mitochondrial ATP6 mRNA; Atp6p encodes a subunit of F1F0 ATP synthase; localized to the mitochondrial inner membrane
RSM22	119	1.844	-1.352	Mitochondrial ribosomal protein of the small subunit; also predicted to be an S-adenosylmethionine-dependent RNA methyltransferase
ATP18	120	1.844	-1.355	Subunit of the mitochondrial F1F0 ATP synthase; F1F0 ATP synthase is a large, evolutionarily conserved enzyme complex required for ATP synthesis; termed subunit I or subunit j; does not correspond to known ATP synthase subunits in other organisms
MRP21	121	1.833	-1.59	Mitochondrial ribosomal protein of the small subunit; MRP21 exhibits genetic interactions with mutations in the COX2 and COX3 mRNA 5'-untranslated leader sequences
ATP23	122	1.817	-0.698	Putative metalloprotease of the mitochondrial inner membrane; required for processing of Atp6p; has an additional role in assembly of the F0 sector of the F1F0 ATP synthase complex; substrate of the Mia40p-Erv1p disulfide relay system, and folding is assisted by Mia40p
MSS116	123	1.788	-1.112	Mitochondrial transcription elongation factor; DEAD-box protein; required for efficient splicing of mitochondrial Group I and II introns; non-polar RNA helicase that also facilities strand annealing; promotes RNA folding by stabilizing an early assembly intermediate
MRPL22	124	1.771	-0.935	Mitochondrial ribosomal protein of the large subunit
MRP7	125	1.769	-1.312	Mitochondrial ribosomal protein of the large subunit
IMP1	126	1.758	-0.702	Catalytic subunit of mitochondrial inner membrane peptidase complex; required for maturation of mitochondrial proteins of the intermembrane space; complex contains two catalytic subunits (Imp1p and Imp2p that differ in substrate specificty) and Som1p
UBX4	127	1.755	-0.772	UBX domain-containing protein that interacts with Cdc48p; involved in degradation of polyubiquitinated proteins via the ERAD (ER-associated degradation) pathway; modulates the Cdc48p-Nplp-Ufd1p AAA ATPase complex during its role in delivery of misfolded proteins to the proteasome; protein abundance increases in response to DNA replication stress
MRPL35	128	1.737	-1.528	Mitochondrial ribosomal protein of the large subunit

RIM101	129	1.734	-0.854	Cys2His2 zinc-finger transcriptional repressor; involved in alkaline responsive gene repression as part of adaptation to alkaline conditions; involved in cell wall assembly; required for alkaline pH-stimulated haploid invasive growth and sporulation; activated by alkaline-dependent proteolytic processing which results in removal of the C-terminal tail; similar to A. nidulans PacC
MRPL4	130	1.728	-1.675	Mitochondrial ribosomal protein of the large subunit; homolog of prokaryotic L29 ribosomal protein; located at the ribosomal tunnel exit
COX20	132	1.703	-0.164	Mitochondrial inner membrane protein; required for proteolytic processing of Cox2p and its assembly into cytochrome c oxidase
COX11	133	1.68	-0.744	Protein required for delivery of copper to Cox1p; mitochondrial inner membrane protein; association with mitochondrial ribosomes suggests that copper delivery may occur during translation of Cox1p
PPA2	134	1.663	-1.252	Mitochondrial inorganic pyrophosphatase; required for mitochondrial function and possibly involved in energy generation from inorganic pyrophosphate; human ortholog, PPA2, functionally complements the null mutant; mutations in human PPA2 cause a mitochondrial disease resulting in sudden unexpected cardiac arrest in infants
MRPL33	135	1.64	-1.337	Mitochondrial ribosomal protein of the large subunit
BUG1	136	1.638	-0.163	Cis-golgi localized protein involved in ER to Golgi transport; forms a complex with the mammalian GRASP65 homolog, Grh1p; mutants are compromised for the fusion of ER-derived vesicles with Golgi membranes
NRP1	137	1.631	-0.873	Putative RNA binding hypothetical protein; localizes to stress granules induced by glucose deprivation; predicted to be involved in ribosome biogenesis
RPO41	138	1.629	-1.276	Mitochondrial RNA polymerase; single subunit enzyme similar to those of T3 and T7 bacteriophages; requires a specificity subunit encoded by MTF1 for promoter recognition; Mtf1p interacts with and stabilizes the Rpo41p-promoter complex, enhancing DNA bending and melting to facilitate pre-initiation open complex formation; Rpo41p also synthesizes RNA primers for mitochondrial DNA replication
YPL183W	140	1.624	-0.494	No Description Found
MRPL15	139	1.624	-1.643	Mitochondrial ribosomal protein of the large subunit
AIM31	141	1.607	-0.271	No Description Found
YMR244C-A	142	1.604	0.675	Protein involved in cytochrome c oxidase (Complex IV) assembly; involved in delivery of copper to Complex IV; also required for efficient formation of respiratory supercomplexes comprised of Complexes III and IV; localizes to the mitochondrial intermembrane space; ortholog implicated in cardiac defects in zebrafish and human; transcription is induced in response to the DNA-damaging agent MMS; protein abundance increases in response to DNA replication stress
MRP10	143	1.555	-1.029	Mitochondrial ribosomal protein of the small subunit; contains twin cysteine-x9-cysteine motifs; oxidized by Mia40p during import into mitochondria
OMS1	144	1.537	-0.217	Protein integral to the mitochondrial membrane; has a conserved methyltransferase motif and is predicted to be an RNA methyltransferase; multicopy suppressor of respiratory defects caused by OXA1 mutations
ATP2	145	1.525	-0.564	Beta subunit of the F1 sector of mitochondrial F1F0 ATP synthase; which is a large, evolutionarily conserved enzyme complex required for ATP synthesis; F1 translationally regulates ATP6 and ATP8 expression to achieve a balanced output of ATP synthase genes encoded in nucleus and mitochondria; phosphorylated
MRPS12	146	1.521	-1.167	Mitochondrial protein; may interact with ribosomes based on co-purification experiments; similar to E. coli and human mitochondrial S12 ribosomal proteins
GDT1	147	1.506	0.085	Calcium transporter localized to the cis- and medial-Golgi apparatus; required for protein glycosylation; GFP-fusion protein localizes to the vacuole; TMEM165, a human gene which causes Congenital Disorders of Glycosylation is orthologous and functionally complements the null allele; expression pattern and physical interactions suggest a possible role in ribosome biogenesis; expression reduced in a gcr1 null mutant
RML2	148	1.482	-1.504	Mitochondrial ribosomal protein of the large subunit (L2); has similarity to E. coli L2 ribosomal protein; mutant allele (fat21) causes inability to utilize oleate, and

				induce oleic acid oxidation; may interfere with activity of the Adr1p transcription factor
MAK10	149	1.481	-0.222	Non-catalytic subunit of the NatC N-terminal acetyltransferase; required for replication of dsRNA virus; expression is glucose-repressible; human NatC ortholog, Naa35, requires co-expression of the human catalytic subunit, Naa30, to functionally complement the null allele
UFO1	150	1.478	-0.485	F-box receptor protein; subunit of the Skp1-Cdc53-F-box receptor (SCF) E3 ubiquitin ligase complex; binds to phosphorylated Ho endonuclease, allowing its ubiquitination by SCF and subsequent degradation
OST4	151	1.476	-1.58	Subunit of the oligosaccharyltransferase complex of the ER lumen; complex catalyzes protein asparagine-linked glycosylation; type I membrane protein required for incorporation of Ost3p or Ost6p into the OST complex
MRPL40	152	1.472	-1.533	Mitochondrial ribosomal protein of the large subunit
RSM19	153	1.418	-1.292	Mitochondrial ribosomal protein of the small subunit; has similarity to E. coli S19 ribosomal protein
MAC1	154	1.417	-0.609	Copper-sensing transcription factor; involved in regulation of genes required for high affinity copper transport; required for regulation of yeast copper genes in response to DNA-damaging agents; undergoes changes in redox state in response to changing levels of copper or MMS
COX16	155	1.399	-1.817	Mitochondrial inner membrane protein; required for assembly of cytochrome c oxidase
PET127	156	1.394	-0.32	Protein with a role in 5'-end processing of mitochondrial RNAs; located in the mitochondrial membrane
VAM3	157	1.389	0.004	Syntaxin-like vacuolar t-SNARE; functions with Vam7p in vacuolar protein trafficking; mediates docking/fusion of late transport intermediates with the vacuole; has an acidic di-leucine sorting signal and C-terminal transmembrane region
YJR120W	158	1.384	-0.499	hypothetical protein; essential for growth under anaerobic conditions; mutation causes decreased expression of ATP2, impaired respiration, defective sterol uptake, and altered levels/localization of ABC transporters Aus1p and Pdr11p
RPN4	159	1.379	-0.327	Transcription factor that stimulates expression of proteasome genes; Rpn4p levels are in turn regulated by the 26S proteasome in a negative feedback control mechanism; RPN4 is transcriptionally regulated by various stress responses; relative distribution to the nucleus increases upon DNA replication stress
ATP12	160	1.378	-0.901	Assembly factor for F1 sector of mitochondrial F1F0 ATP synthase; conserved protein; required for assembly of alpha and beta subunits into F1 sector of mitochondrial F1F0 ATP synthase; human homolog ATPAF2 can complement yeast atp12 mutant; mutation of human homolog reduces active ATP synthase levels and is associated with the disorder ATPAF2 deficiency
MAK3	161	1.367	0.395	Catalytic subunit of the NatC type N-terminal acetyltransferase (NAT); involved in subcellular targeting of select N-terminally acetylated substrates to the Golgi apparatus (Arl3p and Grh1p) and the inner nuclear membrane (Trm1p); required for replication of dsRNA virus; human NatC ortholog, Naa60, functionally complements the null, requiring either auxiliary subunit Mak10p or co-expression of human ortholog, Naa35; Naa60, the human NatF gene, also complements the null allele
SSN3	162	1.361	-0.651	Cyclin-dependent protein kinase; component of RNA polymerase II holoenzyme; involved in phosphorylation of the RNA polymerase II C-terminal domain; involved in glucose repression
CYC2	163	1.337	-0.348	Mitochondrial peripheral inner membrane protein; contains a FAD cofactor in a domain exposed in the intermembrane space; exhibits redox activity in vitro; likely participates in ligation of heme to acytochromes c and c1 (Cyc1p and Cyt1p)
COX8	164	1.336	0.048	Subunit VIII of cytochrome c oxidase (Complex IV); Complex IV is the terminal member of the mitochondrial inner membrane electron transport chain
CRD1	165	1.328	0.085	Cardiolipin synthase; produces cardiolipin, which is a phospholipid of the mitochondrial inner membrane that is required for normal mitochondrial membrane potential and function and for correct integration of membrane-multispanning proteins into the mitochondrial outer membrane; required to

				maintain tubular mitochondrial morphology and functions in mitochondrial fusion; also required for normal vacuolar ion homeostasis
PRY1	166	1.32	0.343	Sterol binding protein involved in the export of acetylated sterols; secreted glycoprotein and member of the CAP protein superfamily (cysteine-rich secretory proteins (CRISP), antigen 5, and pathogenesis related 1 proteins); sterol export function is redundant with that of PRY2; may be involved in detoxification of hydrophobic compounds; PRY1 has a paralog, PRY2, that arose from the whole genome duplication
ATP7	167	1.293	-1.563	Subunit d of the stator stalk of mitochondrial F1F0 ATP synthase; F1F0 ATP synthase is a large, evolutionarily conserved enzyme complex required for ATP synthesis
YER084W	168	1.287	-0.02	hypothetical protein; expressed at both mRNA and protein levels
MRPS8	169	1.283	-1.335	Mitochondrial ribosomal protein of the small subunit
NIP100	170	1.273	-1.064	Large subunit of the dynactin complex; dynactin is involved in partitioning the mitotic spindle between mother and daughter cells; putative ortholog of mammalian p150(glued)
COG1	171	1.27	-1.577	Essential component of the conserved oligomeric Golgi complex; a cytosolic tethering complex (Cog1p through Cog8p) that functions in protein trafficking to mediate fusion of transport vesicles to Golgi compartments
YGL079W	172	1.269	-0.154	Subunit of the BLOC-1 complex involved in endosomal maturation; null mutant is sensitive to drug inducing secretion of vacuolar cargo; GFP-fusion protein localizes to the endosome
ICP55	173	1.267	0.084	Mitochondrial aminopeptidase; cleaves the N termini of at least 38 imported proteins after cleavage by the mitochondrial processing peptidase (MPP), thereby increasing their stability; member of the aminopeptidase P family
YDR493W	174	1.25	-0.135	Protein required for assembly of the cytochrome bc(1) complex; acts as a chaperone for Rip1p and facilitates its insertion into the complex at a late stage of assembly; localized to the mitochondrial matrix; null mutant exhibits a respiratory growth defect and reduced mitochondrial zinc levels, which is characteristic of mutations affecting bc(1) complex assembly; member of the LYR protein family; human LYRM7 is a functional ortholog
MMR1	175	1.238	0.106	Phosphorylated protein of the mitochondrial outer membrane; localizes only to mitochondria of the bud; interacts with Myo2p to mediate mitochondrial distribution to buds; mRNA is targeted to the bud via the transport system involving She2p
CYC1	176	1.236	-0.149	Cytochrome c, isoform 1; also known as iso-1-cytochrome c; electron carrier of mitochondrial intermembrane space that transfers electrons from ubiquinone-cytochrome c oxidoreductase to cytochrome c oxidase during cellular respiration; CYC1 has a paralog, CYC7, that arose from the whole genome duplication; human homolog CYC1 can complement yeast null mutant; mutations in human CYC1 cause insulin-responsive hyperglycemia
GEF1	177	1.235	-0.261	Voltage-gated chloride channel; localized to the golgi, the endosomal system, and plasma membrane; involved in cation homeostasis; highly homologous to vertebrate voltage-gated chloride channels; modulates TBSV model (+) RNA virus replication by regulating copper metabolism
STV1	178	1.226	0.895	Subunit a of the vacuolar-ATPase V0 domain; one of two isoforms (Stv1p and Vph1p); Stv1p is located in V-ATPase complexes of the Golgi and endosomes while Vph1p is located in V-ATPase complexes of the vacuole
YSC83	179	1.213	0.581	Non-essential mitochondrial hypothetical protein; mRNA induced during meiosis, peaking between mid to late prophase of meiosis I; similar to S. douglasii YSD83
MRPL25	180	1.204	-1.518	Mitochondrial ribosomal protein of the large subunit; mutation confers increased replicative lifespan
YML009c	181	1.202	0.069	Mitochondrial ribosomal protein of the large subunit

SUPPLEMENTARY TABLE S6: Bottom Betaxanthin Hits

gene	Yield_Rank	Yield_Score	Size_Score	Description

BUD25	4227	-2.876	-2.56	Protein involved in bud-site selection; diploid mutants display a random budding pattern instead of the wild-type bipolar pattern
THP1	4226	-2.696	-2.586	Nuclear pore-associated protein; component of TREX-2 complex (Sac3p-Thp1p-Sus1p-Cdc31p) involved in transcription elongation and mRNA export from the nucleus; involved in post-transcriptional tethering of active genes to the nuclear periphery and to non-nascent mRNP; contains a PAM domain implicated in protein-protein binding
PRS3	4225	-2.628	-2.227	5-phospho-ribosyl-1(alpha)-pyrophosphate synthetase; synthesizes PRPP, which is required for nucleotide, histidine, and tryptophan biosynthesis; one of five related enzymes, which are active as heteromultimeric complexes
GAL10	4224	-2.588	-1.592	UDP-glucose-4-epimerase; catalyzes interconversion of UDP-galactose and UDP-D-glucose in galactose metabolism; also catalyzes conversion of alpha-D-glucose or alpha-D-galactose to their beta-anomers; human homolog GALE implicated in galactosemia, can complement yeast null mutant
VMA13	4223	-2.561	-1.948	Subunit H of the V1 peripheral membrane domain of V-ATPase; part of the electrogenic proton pump found throughout the endomembrane system; serves as an activator or a structural stabilizer of the V-ATPase; the V1 peripheral membrane domain of the vacuolar H+-ATPase (V-ATPase) has eight subunits
APQ12	4222	-2.405	-2.158	Nuclear envelope/ER integral membrane protein; interacts and functions with Brr6p and Brl1p in lipid homeostasis; mutants are defective in nuclear pore complex biogenesis, nuclear envelope morphology, mRNA export from the nucleus and are sensitive to sterol biosynthesis inhibitors and membrane fluidizing agents; exhibits synthetic lethal genetic interactions with genes involved in lipid metabolism
BUR2	4221	-2.368	-1.14	Cyclin for the Sgv1p (Bur1p) protein kinase; Sgv1p and Bur2p comprise the CDK-cyclin BUR kinase complex which is involved in transcriptional regulation through its phosphorylation of the carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II (Rpo21p); BUR kinase is also involved in the recruitment of Spt6p to the CTD at the onset of transcription
CAX4	4220	-2.348	-2.555	Dolichyl pyrophosphate (Dol-P-P) phosphatase; has a luminally oriented active site in the ER; cleaves the anhydride linkage in Dol-P-P; required for Dol-P-P-linked oligosaccharide intermediate synthesis and protein N-glycosylation
ZUO1	4219	-2.32	-2.31	Ribosome-associated chaperone; zuotin functions in ribosome biogenesis and as a chaperone for nascent polypeptide chains in partnership with Ssz1p and SSb1/2; contains a DnaJ domain and functions as a J-protein partner for Ssb1p and Ssb2p; human gene DNAJC2 can partially complement yeast zuo1 null mutant
VPS34	4218	-2.286	-1.326	Phosphatidylinositol (PI) 3-kinase that synthesizes PI-3-phosphate; forms membrane-associated signal transduction complex with Vps15p to regulate protein sorting; activated by the GTP-bound form of Gpa1p; a fraction is localized, with Vps15p, to nuclear pores at nucleus-vacuole junctions and may facilitate transcription elongation for genes positioned at the nuclear periphery
DBP7	4217	-2.286	-2.711	Putative ATP-dependent RNA helicase of the DEAD-box family; involved in ribosomal biogenesis; required at post-transcriptional step for efficient retrotransposition; essential for growth under anaerobic conditions
MAP1	4216	-2.276	-1.929	Methionine aminopeptidase; catalyzes the cotranslational removal of N-terminal methionine from nascent polypeptides; function is partially redundant with that of Map2p
PDX3	4215	-2.218	-0.592	Pyridoxine (pyridoxamine) phosphate oxidase; has homologs in E. coli and Myxococcus xanthus; transcription is under the general control of nitrogen metabolism
RPS10A	4214	-2.211	-2.328	Protein component of the small (40S) ribosomal subunit; homologous to mammalian ribosomal protein S10, no bacterial homolog; RPS10A has a paralog, RPS10B, that arose from the whole genome duplication; mutations in the human homolog associated with Diamond-Blackfan anemia
ERG3	4213	-2.204	-2.125	C-5 sterol desaturase; glycoprotein that catalyzes the introduction of a C-5(6) double bond into episterol, a precursor in ergosterol biosynthesis; transcriptionally down-regulated when ergosterol is in excess; mutants are viable, but cannot grow on non-fermentable carbon sources; substrate of HRD ubiquitin ligase; mutation is functionally complemented by human SC5D

RTF1	4212	-2.172	-1.319	Subunit of RNAPII-associated chromatin remodeling Paf1 complex; regulates gene expression by directing cotranscriptional histone modification, influences transcription and chromatin structure through several independent functional domains; directly or indirectly regulates DNA-binding properties of Spt15p and relative activities of different TATA elements; involved in transcription elongation as demonstrated by the G-less-based run-on (GLRO) assay
RPL27A	4211	-2.118	-2.196	Ribosomal 60S subunit protein L27A; homologous to mammalian ribosomal protein L27, no bacterial homolog; RPL27A has a paralog, RPL27B, that arose from the whole genome duplication
NUP84	4210	-2.097	-2.371	Subunit of the Nup84p subcomplex of the nuclear pore complex (NPC); contributes to nucleocytoplasmic transport and NPC biogenesis; also plays roles in several processes that may require localization of genes or chromosomes at the nuclear periphery, including double-strand break repair, transcription and chromatin silencing; homologous to human NUP107
YGL081W	4209	-2.095	-1.939	hypothetical protein; non-essential gene; interacts genetically with CHS5, a gene involved in chitin biosynthesis
PRO2	4208	-2.085	-2.059	Gamma-glutamyl phosphate reductase; catalyzes the second step in proline biosynthesis
1.00 BRE	4207	-2.064	-1.146	E3 ubiquitin ligase; forms heterodimer with Rad6p to regulate K63 polyubiquitination in response to oxidative stress and to monoubiquinate histone H2B-K123, which is required for the subsequent methylation of histone H3-K4 and H3-K79; required for DSBR, transcription, silencing, and checkpoint control; interacts with RNA-binding protein Npl3p, linking histone ubiquitination to mRNA processing; Bre1p-dependent histone ubiquitination promotes pre-mRNA splicing
YDL119C	4206	-2.063	-0.958	Mitochondrial glycine transporter; required for the transport of glycine into mitochondria for initiation of heme biosynthesis, with YMC1 acting as a secondary transporter; homolog of human SLC25A38, a mitochondrial glycine transporter associated with nonsyndromic autosomal recessive congenital sideroblastic anemia; human SLC25A38 can complement the heme deficiency associated with the null mutant; GFP-fusion protein is induced in response to the DNA-damaging agent MMS
SLM2	852	-2.058	-2.379	Phosphoinositide PI4,5P(2) binding protein, forms a complex with Slm1p; acts downstream of Mss4p in a pathway regulating actin cytoskeleton organization in response to stress; TORC2 complex substrate and effector; SLM2 has a paralog, SLM1, that arose from the whole genome duplication
VMA2	4204	-2.054	-1.902	Subunit B of V1 peripheral membrane domain of vacuolar H+-ATPase; electrogenic proton pump found throughout the endomembrane system; contains nucleotide binding sites; also detected in the cytoplasm; protein abundance increases in response to DNA replication stress; human homolog ATP6V1B1, implicated in autosomal-recessive distal renal tubular acidosis (RTA) with sensorineural deafness, complements yeast null mutant
ANP1	4203	-2.051	-1.709	Subunit of the alpha-1,6 mannosyltransferase complex; type II membrane protein; has a role in retention of glycosyltransferases in the Golgi; involved in osmotic sensitivity and resistance to aminonitrophenyl propanediol
RRN10	4202	-2.045	-1.957	Protein involved in promoting high level transcription of rDNA; subunit of UAF (upstream activation factor) for RNA polymerase I
SLM4	4201	-2.044	-2.082	Component of the EGO and GSE complexes; essential for integrity and function of EGO; EGO is involved in the regulation of microautophagy and GSE is required for proper sorting of amino acid permease Gap1p; gene exhibits synthetic genetic interaction with MSS4
SFP1	4200	-2.031	-1.974	Regulates transcription of ribosomal protein and biogenesis genes; regulates response to nutrients and stress, G2/M transitions during mitotic cell cycle and DNA-damage response, and modulates cell size; regulated by TORC1 and Mrs6p; sequence of zinc finger, ChIP localization data, and protein-binding microarray (PBM) data, and computational analyses suggest it binds DNA directly at highly active RP genes and indirectly through Rap1p at others; can form the [ISP+] prion
RPS19B	4199	-2.023	-2.396	Protein component of the small (40S) ribosomal subunit; required for assembly and maturation of pre-40 S particles; homologous to mammalian ribosomal protein S19, no bacterial homolog; mutations in human RPS19 are associated with Diamond Blackfan anemia; RPS19B has a paralog, RPS19A, that arose from the whole genome duplication

PHO88	4198	-2.021	-1.957	Protein involved in SRP-independent targeting of substrates to the ER; component of an alternative ER targeting pathway that has partial functional redundancy with the GET pathway; preference for substrates with downstream transmembrane domains; interacts with Snd1p, Env10p/Snd2p, and Sec61p-translocon subunits; can compensate for loss of SRP; role in phosphate transport, interacting with pho88, and in the maturation of secretory proteins
KGD2	4196	-2.013	-0.276	Dihydrolipoyl transsuccinylase; component of the mitochondrial alpha-ketoglutarate dehydrogenase complex, which catalyzes the oxidative decarboxylation of alpha-ketoglutarate to succinyl-CoA in the TCA cycle; phosphorylated
UAF30	4197	-2.013	-2.153	Subunit of UAF (upstream activation factor) complex; UAF is an RNA polymerase I specific transcription stimulatory factor composed of Uaf30p, Rrn5p, Rrn9p, Rrn10p, histones H3 and H4; targeting factor for the UAF that facilitates activation of many rDNA genes; deletion decreases cellular growth rate; UAF30 has a paralog, TRI1, that arose from the whole genome duplication
GON7	4195	-2.006	-2.108	Component of the EKC/KEOPS protein complex; EKC/KEOPS complex is required for t6A tRNA modification and telomeric TG1-3 recombination; may have role in transcription; implicated in osmotic stress response; other complex members are Kae1p, Cgi121p, Pcc1p, and Bud32p
YHR039C-B	4194	-1.994	-2.111	No Description Found
VMA22	4193	-1.99	-1.698	Protein that is required for vacuolar H+-ATPase (V-ATPase) function; peripheral membrane protein; not an actual component of the V-ATPase complex; functions in the assembly of the V-ATPase; localized to the yeast endoplasmic reticulum (ER)
PEP12	4192	-1.982	-2.276	Target membrane receptor (t-SNARE); for vesicular intermediates traveling between the Golgi apparatus and the vacuole; controls entry of biosynthetic, endocytic, and retrograde traffic into the prevacuolar compartment; syntaxin
BUD31	4191	-1.974	-2.479	Component of the SF3b subcomplex of the U2 snRNP; increases efficiency of first and second step pre-mRNA splicing; diploid mutants display a random budding pattern instead of the wild-type bipolar pattern; facilitates passage through G1/S Start, but is not required for G2/M transition or exit from mitosis
SNF7	4190	-1.973	-0.215	One of four subunits of the ESCRT-III complex; involved in the sorting of transmembrane proteins into the multivesicular body (MVB) pathway; recruited from the cytoplasm to endosomal membranes; ESCRT-III stands for endosomal sorting complex required for transport III
BDF1	4189	-1.964	-2.019	Protein involved in transcription initiation; functions at TATA-containing promoters; associates with the basal transcription factor TFIID; contains two bromodomains; corresponds to the C-terminal region of mammalian TAF1; redundant with Bdf2p; BDF1 has a paralog, BDF2, that arose from the whole genome duplication
PEX32	4188	-1.963	-1.666	Peroxisomal integral membrane protein; involved in negative regulation of peroxisome size; partially functionally redundant with Pex31p; genetic interactions suggest action at a step downstream of steps mediated by Pex28p and Pex29p
RPS17A	4187	-1.952	-2.668	Ribosomal protein 51 (rp51) of the small (40s) subunit; homologous to mammalian ribosomal protein S17, no bacterial homolog; RPS17A has a paralog, RPS17B, that arose from the whole genome duplication
GAL4	4186	-1.949	-1.9	DNA-binding transcription factor required for activating GAL genes; responds to galactose; repressed by Gal80p and activated by Gal3p
INO4	4185	-1.947	-2.31	Transcription factor involved in phospholipid synthesis; required for derepression of inositol-choline-regulated genes involved in phospholipid synthesis; forms a complex, with Ino2p, that binds the inositol-choline-responsive element through a basic helix-loop-helix domain
VMA6	4184	-1.938	-2.338	Subunit d of the V0 integral membrane domain of V-ATPase; part of the electrogenic proton pump found in the endomembrane system; required for V1 domain assembly on the vacuolar membrane; the V0 integral membrane domain of vacuolar H+-ATPase (V-ATPase) has five subunits
SNX4	4183	-1.929	-1.286	Sorting nexin; involved in retrieval of late-Golgi SNAREs from post-Golgi endosomes to the trans-Golgi network and in cytoplasm to vacuole transport; contains a PX phosphoinositide-binding domain; forms complexes with Snx41p and with Atg20p

THR1	4182	-1.908	-1.864	Homoserine kinase; conserved protein required for threonine biosynthesis; long-lived protein that is preferentially retained in mother cells and forms cytoplasmic filaments; expression is regulated by the GCN4-mediated general amino acid control pathway
ILV1	4181	-1.902	-2.294	Threonine deaminase, catalyzes first step in isoleucine biosynthesis; expression is under general amino acid control; ILV1 locus exhibits highly positioned nucleosomes whose organization is independent of known ILV1 regulation
PUT2	4180	-1.888	0.043	Delta-1-pyrroline-5-carboxylate dehydrogenase; nuclear-encoded mitochondrial protein involved in utilization of proline as sole nitrogen source; deficiency of human homolog ALDH4A1 causes type II hyperprolinemia (HPII), an autosomal recessive inborn error of metabolism; human homolog ALDH4A1 can complement yeast null mutant
JJJ1	4179	-1.88	-1.67	Co-chaperone that stimulates the ATPase activity of Ssa1p; required for a late step of ribosome biogenesis; associated with the cytosolic large ribosomal subunit; contains a J-domain; mutation causes defects in fluid-phase endocytosis
RPL19B	4178	-1.872	-1.795	Ribosomal 60S subunit protein L19B; rpl19a and rpl19b single null mutations result in slow growth, while the double null mutation is lethal; homologous to mammalian ribosomal protein L19, no bacterial homolog; RPL19B has a paralog, RPL19A, that arose from the whole genome duplication
SKI7	4177	-1.854	-2.449	GTP-binding protein that couples the Ski complex and exosome; putative pseudo-translational GTPase involved in 3'-to-5' mRNA decay pathway; interacts with both the cytoplasmic exosome and the Ski complex; eRF3-like domain targets nonstop mRNA for degradation; null mutants have a superkiller phenotype; SKI7 has a paralog, HBS1, that arose from the whole genome duplication
SNT309	4176	-1.852	-0.812	Member of the NineTeen Complex (NTC); this complex contains Prp19p and stabilizes U6 snRNA in catalytic forms of the spliceosome containing U2, U5, and U6 snRNAs; interacts physically and genetically with Prp19p
PPA1	4175	-1.836	-1.962	No Description Found
SAP30	4174	-1.834	-0.847	Component of Rpd3L histone deacetylase complex; involved in silencing at telomeres, rDNA, and silent mating-type loci; involved in telomere maintenance
VPS25	4173	-1.829	-0.513	Component of the ESCRT-II complex; ESCRT-II is involved in ubiquitin-dependent sorting of proteins into the endosome
VMA5	4172	-1.828	-1.91	Subunit C of the V1 peripheral membrane domain of V-ATPase; part of the electrogenic proton pump found throughout the endomembrane system; required for the V1 domain to assemble onto the vacuolar membrane; the V1 peripheral membrane domain of vacuolar H+-ATPase (V-ATPase) has eight subunits
ARO1	4171	-1.825	-0.798	Pentafunctional arom protein; catalyzes steps 2 through 6 in the biosynthesis of chorismate, which is a precursor to aromatic amino acids
1.00 SHP	4170	-1.809	-2.224	UBX domain-containing substrate adaptor for Cdc48p; ubiquitin regulatory X domain-containing protein that acts as a substrate recruiting cofactor for Cdc48p; positively regulates Glc7p PPase activity to promote growth and mitotic progression in complex with Cdc48p; ubiquitinated protein interactor involved in ER-associated degradation (ERAD); regulated by nuclear Ub-dependent degradation (INMAD pathway) independent of the Asi and Doa10 complexes; homolog of human p47 (NSFL1C)
GCN4	4169	-1.799	-1.841	bZIP transcriptional activator of amino acid biosynthetic genes; activator responds to amino acid starvation; expression is tightly regulated at both the transcriptional and translational levels
RPL14A	4168	-1.792	-1.92	Ribosomal 60S subunit protein L14A; N-terminally acetylated; homologous to mammalian ribosomal protein L14, no bacterial homolog; RPL14A has a paralog, RPL14B, that arose from the whole genome duplication
ADE12	4167	-1.792	-2.008	Adenylosuccinate synthase; catalyzes the first step in synthesis of adenosine monophosphate from inosine 5'monophosphate during purine nucleotide biosynthesis; exhibits binding to single-stranded autonomously replicating (ARS) core sequence
YDJ1	4166	-1.787	-2.33	Type I HSP40 co-chaperone; involved in regulation of HSP90 and HSP70 functions; acts as an adaptor that helps Rsp5p recognize cytosolic misfolded proteins for ubiquitination after heat shock; critical for determining cell size at Start as a function of growth rate; involved in protein translocation across membranes;

				member of the DnaJ family; chimeric protein in which human p58IPK J domain replaces yeast Ydj1p J domain can complement yeast ydj1 mutant
RPL13B	4164	-1.785	-2.465	Ribosomal 60S subunit protein L13B; not essential for viability; homologous to mammalian ribosomal protein L13, no bacterial homolog; RPL13B has a paralog, RPL13A, that arose from the whole genome duplication
NUP133	4165	-1.785	-1.333	Subunit of Nup84p subcomplex of nuclear pore complex (NPC); contributes to nucleocytoplasmic transport, NPC biogenesis; is involved in establishment of a normal nucleocytoplasmic concentration gradient of GTPase Gsp1p; also plays roles in several processes that may require localization of genes or chromosomes at nuclear periphery, including double-strand break repair, transcription and chromatin silencing; relocalizes to cytosol in response to hypoxia; homolog of human NUP133
ZAP1	4163	-1.778	-2.274	Zinc-regulated transcription factor; binds to zinc-responsive promoters to induce transcription of certain genes in presence of zinc, represses other genes in low zinc; regulates its own transcription; contains seven zinc-finger domains
ERG2	4162	-1.773	-1.505	C-8 sterol isomerase; catalyzes isomerization of delta-8 double bond to delta-7 position at an intermediate step in ergosterol biosynthesis; transcriptionally down-regulated when ergosterol is in excess; mutation is functionally complemented by human EBP
VPS20	4161	-1.77	-0.544	Myristoylated subunit of the ESCRT-III complex; the endosomal sorting complex required for transport of transmembrane proteins into the multivesicular body pathway to the lysosomal/vacuolar lumen; cytoplasmic protein recruited to endosomal membranes
LST7	4160	-1.755	-0.978	Subunit of the Lst4p-Lst7p GTPase activating protein complex for Gtr2p; stimulates the GTPase activity of Rag family GTPase Gtr2p, within the context of the Gtr1p-Gtr2p heterodimer, after amino acid stimulation; required for activation of TORC1 in response to amino acid stimulation; recruited to the vacuolar membrane during amino acid starvation and released from the membrane by TORC1; required for the transport of amino acid permease Gap1p from the Golgi to the cell surface
RNR1	4159	-1.753	-1.804	Major isoform of large subunit of ribonucleotide-diphosphate reductase; the RNR complex catalyzes rate-limiting step in dNTP synthesis, regulated by DNA replication and DNA damage checkpoint pathways via localization of small subunits; relative distribution to the nucleus increases upon DNA replication stress; RNR1 has a paralog, RNR3, that arose from the whole genome duplication
VPS9	4158	-1.738	-1.351	Guanine nucleotide exchange factor (GEF) and ubiquitin receptor; involved in vesicle-mediated vacuolar transport, including Golgi-endosome trafficking and sorting through the multivesicular body (MVB); stimulates the intrinsic guanine nucleotide exchange activity of Rab family members (Vps21p/Ypt52p/Ypt53p); partially redundant with GEF MUK1; required for localization of the CORVET complex to endosomes; similar to mammalian ras inhibitors; contains a Ub-interacting CUE domain
ASF1	4157	-1.733	-1.688	Nucleosome assembly factor; involved in chromatin assembly, disassembly; required for recovery after DSB repair; role in H3K56 acetylation required for expression homeostasis, buffering mRNA synthesis rate against gene dosage changes in S phase; anti-silencing protein, derepresses silent loci when overexpressed; role in regulating Ty1 transposition; relocalizes to cytosol under hypoxia; growth defect of asf1 null is functionally complemented by either human ASF1A or ASF1B
MTD1	4156	-1.73	-0.674	NAD-dependent 5,10-methylenetetrahydrafolate dehydrogenase; plays a catalytic role in oxidation of cytoplasmic one-carbon units; expression is regulated by Bas1p and Bas2p, repressed by adenine, and may be induced by inositol and choline
REF2	4155	-1.726	-2.698	RNA-binding protein; involved in the cleavage step of mRNA 3'-end formation prior to polyadenylation, and in snoRNA maturation; part of holo-CPF subcomplex APT, which associates with 3'-ends of snoRNA- and mRNA-encoding genes; putative regulatory subunit of type 1 protein phosphatase Glc7p, required for actomyosin ring formation, and for timely dephosphorylation and release of Bnr1p from the division site; relocalizes to the cytosol in response to hypoxia
5.00 CUP	4154	-1.723	-1.837	No Description Found

RPS27B	4153	-1.721	-0.953	Protein component of the small (40S) ribosomal subunit; homologous to mammalian ribosomal protein S27, no bacterial homolog; RPS27B has a paralog, RPS27A, that arose from the whole genome duplication
SSE1	4152	-1.715	-2.065	ATPase component of heat shock protein Hsp90 chaperone complex; serves as nucleotide exchange factor to load ATP onto the SSA class of cytosolic Hsp70s; plays a role in prion propagation and determining prion variants; binds unfolded proteins; member of Hsp110 subclass of HSP70 proteins; deletion results in spindle elongation in S phase; SSE1 has a paralog, SSE2, that arose from the whole genome duplication
MRE11	4151	-1.702	-1.591	Nuclease subunit of the MRX complex with Rad50p and Xrs2p; complex functions in repair of DNA double-strand breaks and in telomere stability; Mre11p associates with Ser/Thr-rich ORFs in premeiotic phase; nuclease activity required for MRX function; widely conserved; forms nuclear foci upon DNA replication stress
VPS28	4150	-1.694	-0.582	Component of the ESCRT-I complex; complex is involved in ubiquitin-dependent sorting of proteins into the endosome; conserved C-terminal domain interacts with ESCRT-III subunit Vps20p; other members include Stp22p, Srn2p, Vps28p, and Mvb12p
MOT2	4149	-1.687	-1.793	Ubiquitin-protein ligase subunit of the CCR4-NOT complex; with Ubc4p, ubiquitinates nascent polypeptide-associated complex subunits and histone demethyase Jhd2p; CCR4-NOT has roles in transcription regulation, mRNA degradation, and post-transcriptional modifications; regulates levels of DNA Polymerase-{alpha} to promote efficient and accurate DNA replication
IBA57	4148	-1.677	-2.133	Protein involved in incorporating iron-sulfur clusters into proteins; mitochondrial matrix protein; involved in the incorporation of iron-sulfur clusters into mitochondrial aconitase-type proteins; activates the radical-SAM family members Bio2p and Lip5p; interacts with Ccr4p in the two-hybrid system
LTV1	4147	-1.676	-2.422	Component of the GSE complex; GSE is required for proper sorting of amino acid permease Gap1p; required for ribosomal small subunit export from nucleus; required for growth at low temperature
SPT4	4146	-1.672	-1.925	Spt4p/5p (DSIF) transcription elongation factor complex subunit; the Spt4/5 complex binds to ssRNA in a sequence-specific manner, and along with RNAP I and II has multiple roles regulating transcriptional elongation, RNA processing, quality control, and transcription-coupled repair; localizes to kinetochores and heterochromatin, influencing chromosomal dynamics and silencing; required for transcription through long trinucleotide repeats in ORFs and non-protein coding regions
KGD1	4145	-1.67	-1.184	Subunit of the mitochondrial alpha-ketoglutarate dehydrogenase complex; catalyzes a key step in the tricarboxylic acid (TCA) cycle, the oxidative decarboxylation of alpha-ketoglutarate to form succinyl-CoA
RPS16B	4144	-1.666	-1.526	Protein component of the small (40S) ribosomal subunit; homologous to mammalian ribosomal protein S16 and bacterial S9; RPS16B has a paralog, RPS16A, that arose from the whole genome duplication
YGL138C	4143	-1.664	-0.92	hypothetical protein; has no significant sequence similarity to any known protein
PLC1	4142	-1.655	-1.107	Phospholipase C; hydrolyzes phosphatidylinositol 4,5-biphosphate (PIP2) to generate the signaling molecules inositol 1,4,5-triphosphate (IP3) and 1,2-diacylglycerol (DAG); involved in regulating many cellular processes; Plc1p and inositol polyphosphates are required for acetyl-CoA homeostasis which regulates global histone acetylation
MRS1	4141	-1.652	-1.61	Splicing protein; required for splicing of two mitochondrial group I introns (BI3 in COB and AI5beta in COX1); forms a splicing complex, containing four subunits of Mrs1p and two subunits of the BI3-encoded maturase, that binds to the BI3 RNA; MRS1 has a paralog, CCE1, that arose from the whole genome duplication
RPL39	4140	-1.637	-2.754	Ribosomal 60S subunit protein L39; required for ribosome biogenesis; loss of both Rpl31p and Rpl39p confers lethality; also exhibits genetic interactions with SIS1 and PAB1; homologous to mammalian ribosomal protein L39, no bacterial homolog
CTK1	4139	-1.634	-1.8	Catalytic (alpha) subunit of C-terminal domain kinase I (CTDK-I); phosphorylates both RNA pol II subunit Rpo21p to affect transcription and pre-mRNA 3' end processing, and ribosomal protein Rps2p to increase translational fidelity; required for H3K36 trimethylation but not dimethylation by Set2p; suggested stimulatory

				role in 80S formation during translation initiation; similar to the Drosophila dCDK12 and human CDK12 and probably CDK13
LHS1	4138	-1.634	-2.372	Molecular chaperone of the endoplasmic reticulum lumen; involved in polypeptide translocation and folding; nucleotide exchange factor for the ER lumenal Hsp70 chaperone Kar2p; regulated by the unfolded protein response pathway
RNR4	4137	-1.633	-2.334	Ribonucleotide-diphosphate reductase (RNR) small subunit; the RNR complex catalyzes the rate-limiting step in dNTP synthesis and is regulated by DNA replication and DNA damage checkpoint pathways via localization of the small subunits; relocalizes from nucleus to cytoplasm upon DNA replication stress; RNR4 has a paralog, RNR2, that arose from the whole genome duplication
GTR1	4136	-1.625	-1.611	Subunit of a TORC1-stimulating GTPase complex; subunit of the heterodimeric Gtr1-Gtr2 GTPase complex that stimulates TORC1 in response to amino acid stimulation; tethered to the vacuolar membrane as part of the EGOC, a complex required for sorting of Gap1p and microautophagy; involved in phosphate transport and telomeric chromatin silencing; activated by the the Iml1p (GAP) subunit of the SEACIT complex; similar to human RagA and RagB
MRF1	4135	-1.605	-0.852	Mitochondrial translation release factor; involved in stop codon recognition and hydrolysis of the peptidyl-tRNA bond during mitochondrial translation; lack of MRF1 causes mitochondrial genome instability
GAC1	4134	-1.6	-0.035	Regulatory subunit for Glc7p type-1 protein phosphatase (PP1); tethers Glc7p to Gsy2p glycogen synthase, binds Hsf1p heat shock transcription factor, required for induction of some HSF-regulated genes under heat shock; GAC1 has a paralog, PIG1, that arose from the whole genome duplication
VPH2	4133	-1.6	-2.136	Integral membrane protein required for V-ATPase function; not an actual component of the vacuolar H+-ATPase (V-ATPase) complex; functions in the assembly of the V-ATPase; localized to the endoplasmic reticulum (ER); involved in methionine restriction extension of chronological lifespan in an autophagy-dependent manner
CDC40	4131	-1.597	-1.031	Pre-mRNA splicing factor; important for catalytic step II of pre-mRNA splicing and plays a role in cell cycle progression, particularly at the G1/S phase transition; required for DNA synthesis during mitosis and meiosis; has WD repeats; thermosensitivity of the cdc40 null mutant is functionally complemented by a chimeric construct containing the N-terminal 156 amino acids of yeast Cdc40p fused to the C-terminal two thirds (297 amino acids) of human CDC40
RAI1	4132	-1.597	-2.355	Nuclear protein with decapping endonuclease activity; targets mRNAs with unmethylated 7-methylguanosine cap structures and 5'-triphosphates; binds to and stabilizes the exoribonuclease Rat1p; required for pre-rRNA processing; relocalizes to the cytosol in response to hypoxia; homologous to human DOM3Z
SSZ1	4130	-1.596	-2.541	Hsp70 protein that interacts with Zuo1p (a DnaJ homolog); interacts with Zuo1p to form a ribosome-associated complex that binds the ribosome via the Zuo1p subunit; also involved in pleiotropic drug resistance via sequential activation of PDR1 and PDR5; binds ATP
CTF4	4129	-1.592	-0.75	Chromatin-associated protein; required for sister chromatid cohesion; interacts with DNA polymerase alpha (Pol1p) and may link DNA synthesis to sister chromatid cohesion
SPT10	4128	-1.584	-2.084	Histone H3 acetylase with a role in transcriptional regulation; sequence-specific activator of histone genes, binds specifically and cooperatively to pairs of UAS elements in core histone promoters, functions at or near TATA box; involved in S phase-specific acetylation of H3K56 at histone promoters, which is required for recruitment of SWI/SNF nucleosome remodeling complex and subsequent transcription
VPS1	4127	-1.582	-1.938	Dynamin-like GTPase required for vacuolar sorting; also involved in actin cytoskeleton organization, endocytosis, late Golgi-retention of some proteins, regulation of peroxisome biogenesis
NPL3	4125	-1.581	-2.174	RNA-binding protein; promotes elongation, regulates termination, and carries poly(A) mRNA from nucleus to cytoplasm; represses translation initiation by binding eIF4G; required for pre-mRNA splicing; interacts with E3 ubiquitin ligase Bre1p, linking histone ubiquitination to mRNA processing; may have role in telomere maintenance; dissociation from mRNAs promoted by Mtr10p; phosphorylated by Sky1p in cytoplasm; protein abundance increases in response to DNA replication stress

1.00 TOP	4126	-1.581	-0.219	Topoisomerase I; nuclear enzyme that relieves torsional strain in DNA by cleaving and re-sealing the phosphodiester backbone; relaxes both positively and negatively supercoiled DNA; functions in replication, transcription, and recombination; role in processing ribonucleoside monophosphates in genomic DNA into irreversible single-strand breaks; enzymatic activity and interaction with Nsr1p are negatively regulated by polyphosphorylation
RPS21A	4124	-1.58	-0.981	Protein component of the small (40S) ribosomal subunit; homologous to mammalian ribosomal protein S21, no bacterial homolog; RPS21A has a paralog, RPS21B, that arose from the whole genome duplication
RPS7A	4123	-1.574	-1.629	Protein component of the small (40S) ribosomal subunit; interacts with Kti11p; deletion causes hypersensitivity to zymocin; homologous to mammalian ribosomal protein S7, no bacterial homolog; RPS7A has a paralog, RPS7B, that arose from the whole genome duplication
MOG1	4122	-1.57	-1.846	Conserved nuclear protein that interacts with GTP-Gsp1p; stimulates nucleotide release from Gsp1p; involved in nuclear protein import; nucleotide release is inhibited by Yrb1p
NSR1	4121	-1.569	-1.414	Nucleolar protein that binds nuclear localization sequences; required for pre-rRNA processing and ribosome biogenesis; binds to single stranded telomeric DNA and mRNA; methylated by Hmt1p; interaction with Top1p and nucleolar localization are negatively regulated by polyphosphorylation
SEC72	4119	-1.545	-1.453	Non-essential subunit of Sec63 complex; with Sec61 complex, Kar2p/BiP and Lhs1p forms a channel competent for SRP-dependent and post-translational SRP-independent protein targeting and import into the ER; other members are Sec63p, Sec62p, and Sec66p
CSF1	4120	-1.545	-1.508	Protein required for fermentation at low temperature; plays a role in the maturation of secretory proteins; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies
UGO1	4118	-1.537	-2.077	Outer membrane component of the mitochondrial fusion machinery; binds to Fzo1p and Mgm1p to link these two GTPases during mitochondrial fusion; involved in fusion of both the outer and inner membranes; facilitates dimerization of Fzo1p during fusion; import into the outer membrane is mediated by Tom70p and Mim1p; has similarity to carrier proteins but likely not a transporter; similar to human SLC25A46 implicated in optic atroprophy spectrum disorder
PIH1	4117	-1.518	-0.808	Component of the conserved R2TP complex (Rvb1-Rvb2-Tah1-Pih1); R2TP complex interacts with Hsp90 (Hsp82p and Hsc82p) to mediate assembly large protein complexes such as box C/D snoRNPs and RNA polymerase II
RUP1	4116	-1.513	0.084	Protein that regulates ubiquitination of Rsp5p; has a WW domain consensus motif of PPPSY (residues 131-135) that mediates binding of Rsp5p to Ubp2p; contains an UBA domain; relative distribution to the nucleus increases upon DNA replication stress
PTK2	4115	-1.502	-0.646	Serine/threonine protein kinase; involved in regulation of ion transport across plasma membrane; carboxyl terminus is essential for glucose-dependent Pma1p activation via phosphorylation of Pma1p-Ser899; enhances spermine uptake; PTK2 has a paralog, PTK1, that arose from the whole genome duplication
FCY22	4114	-1.479	-0.978	Putative purine-cytosine permease; very similar to Fcy2p but cannot substitute for its function
GND1	4113	-1.473	-1.561	6-phosphogluconate dehydrogenase (decarboxylating); catalyzes an NADPH regenerating reaction in the pentose phosphate pathway; required for growth on D-glucono-delta-lactone and adaptation to oxidative stress; GND1 has a paralog, GND2, that arose from the whole genome duplication
IES6	4112	-1.471	-2.0	Component of the INO80 chromatin remodeling complex; critical for INO80 function; involved in regulation of chromosome segregation and maintenance of normal centromeric chromatin structure; human ortholog INO80C is a member of the human INO80 complex; implicated in DNA repair based on genetic interactions with RAD52 epistasis genes
RPL42A	4111	-1.461	0.215	Ribosomal 60S subunit protein L42A; homologous to mammalian ribosomal protein L36A, no bacterial homolog; RPL42A has a paralog, RPL42B, that arose from the whole genome duplication
RPS6A	4110	-1.456	-1.594	Protein component of the small (40S) ribosomal subunit; homologous to mammalian ribosomal protein S6, no bacterial homolog; phosphorylated on S233

				by Ypk3p in a TORC1-dependent manner, and on S232 in a TORC1/2-dependent manner by Ypk1/2/3p; RPS6A has a paralog, RPS6B, that arose from the whole genome duplication
VMA7	4109	-1.453	-2.325	Subunit F of the V1 peripheral membrane domain of V-ATPase; part of the electrogenic proton pump found throughout the endomembrane system; required for the V1 domain to assemble onto the vacuolar membrane; the V1 peripheral membrane domain of vacuolar H+-ATPase (V-ATPase) has eight subunits
SUV3	4108	-1.442	-1.152	ATP-dependent RNA helicase; component of the mitochondrial degradosome along with the RNase Dss1p; the degradosome associates with the ribosome and mediates RNA turnover; also required during splicing of the COX1 AI5_beta intron; expression of a processed form of human homolog SUPV3L1 carrying an N-terminal deletion of 46 amino acids rescues yeast suv3 null mutant
THR4	4107	-1.44	-1.301	Threonine synthase; conserved protein that catalyzes formation of threonine from O-phosphohomoserine; expression is regulated by the GCN4-mediated general amino acid control pathway
RPL12B	4106	-1.436	-1.981	Ribosomal 60S subunit protein L12B; rpl12a rpl12b double mutant exhibits slow growth and slow translation; homologous to mammalian ribosomal protein L12 and bacterial L11; RPL12B has a paralog, RPL12A, that arose from the whole genome duplication
BUD27	4105	-1.427	-1.606	Unconventional prefoldin protein involved in translation initiation; required for correct assembly of RNAP I, II, and III in an Rpb5p-dependent manner; shuttles between nucleus and cytoplasm; mutants have inappropriate expression of nutrient sensitive genes due to translational derepression of Gcn4p transcription factor; diploid mutants show random budding; ortholog of human URI/RMP
EOS1	4103	-1.425	-1.297	Protein involved in N-glycosylation; deletion mutation confers sensitivity to exidative stress and shows synthetic lethality with mutations in the spindle checkpoint genes BUB3 and MAD1; YNL080C is not an essential gene
SWD1	4104	-1.425	-0.695	Subunit of the COMPASS (Set1C) complex; COMPASS methylates histone H3 on lysine 4 and is required in transcriptional silencing near telomeres; WD40 beta propeller superfamily member with similarity to mammalian Rbbp7
FYV6	4102	-1.42	-2.843	hypothetical protein; required for survival upon exposure to K1 killer toxin; proposed to regulate double-strand break repair via non-homologous end-joining
SOD1	4101	-1.414	-0.707	Cytosolic copper-zinc superoxide dismutase; detoxifies superoxide; stabilizes Yck1p and Yck2p kinases in glucose to repress respiration; phosphorylated by Dun1p, enters nucleus under oxidative stress to promote transcription of stress response genes; human ortholog SOD1 implicated in ALS complements a null allele; abundance increases under DNA replication stress and during exposure to boric acid; localization to mitochondrial intermembrane space is modulated by MICOS complex
PRS5	4100	-1.41	-1.159	5-phospho-ribosyl-1(alpha)-pyrophosphate synthetase; synthesizes PRPP, which is required for nucleotide, histidine, and tryptophan biosynthesis; one of five related enzymes, which are active as heteromultimeric complexes; forms cytoplasmic foci upon DNA replication stress
RAD6	4099	-1.404	-1.666	Ubiquitin-conjugating enzyme (E2); involved in postreplication repair as a heterodimer with Rad18p, regulation of K63 polyubiquitination in response to oxidative stress, DSBR and checkpoint control as a heterodimer with Bre1p, ubiquitin-mediated N-end rule protein degradation as a heterodimer with Ubr1p, ERAD with Ubr1p in the absence of canonical ER membrane ligases, and Rpn4p turnover as part of proteasome homeostasis, in complex with Ubr2p and Mub1p
HOM3	4098	-1.4	-0.607	Aspartate kinase (L-aspartate 4-P-transferase); cytoplasmic enzyme that catalyzes the first step in the common pathway for methionine and threonine biosynthesis; expression regulated by Gcn4p and the general control of amino acid synthesis
GIM5	4097	-1.381	-1.735	Subunit of the heterohexameric cochaperone prefoldin complex; prefoldin binds specifically to cytosolic chaperonin and transfers target proteins to it; prefoldin complex also localizes to chromatin of actively transcribed genes in the nucleus and facilitates transcriptional elongation
HIT1	4096	-1.37	-1.801	Protein involved in C/D snoRNP assembly; regulates abundance of Rsa1p; required for growth at high temperature; similar to human ZNHIT3
RPL31A	4095	-1.369	-2.265	Ribosomal 60S subunit protein L31A; associates with karyopherin Sxm1p; loss of both Rpl31p and Rpl39p confers lethality; homologous to mammalian ribosomal

				protein L31, no bacterial homolog; RPL31A has a paralog, RPL31B, that arose from the whole genome duplication
ELM1	4094	-1.366	-0.565	Serine/threonine protein kinase; regulates the orientation checkpoint, the morphogenesis checkpoint and the metabolic switch from fermentative to oxidative metabolism by phosphorylating the activation loop of Kin4p, Hsl1p and Snf4p respectively; cooperates with Hsl7p in recruiting Hsl1p to the septin ring, a prerequisite for subsequent recruitment, phosphorylation, and degradation of Swe1p; forms part of the bud neck ring; regulates cytokinesis
DOA4	4093	-1.362	-0.211	Ubiquitin hydrolase; deubiquitinates intralumenal vesicle (ILVs) cargo proteins; required for recycling ubiquitin from proteasome-bound ubiquitinated intermediates, acts at the late endosome/prevacuolar compartment to recover ubiquitin from ubiquitinated membrane proteins destined for the vacuole; DOA4 has a paralog, UBP5, that arose from the whole genome duplication
RPL37A	4092	-1.361	-1.766	Ribosomal 60S subunit protein L37A; required for processing of 27SB pre-rRNA and formation of stable 66S assembly intermediates; homologous to mammalian ribosomal protein L37, no bacterial homolog; RPL37A has a paralog, RPL37B, that arose from the whole genome duplication
AIR1	4091	-1.358	-0.176	Zinc knuckle protein; involved in nuclear RNA processing and degradation as a component of the TRAMP complex; stimulates the poly(A) polymerase activity of Pap2p in vitro; AIR1 has a paralog, AIR2, that arose from the whole genome duplication; although Air1p and Air2p are homologous TRAMP subunits, they have nonredundant roles in regulation of substrate specificity of the exosome
YDL133W	4090	-1.356	0.026	Regulator of phospholipase D (Spo14p); interacts with Spo14p and regulates its catalytic activity; capable of buffering the toxicity of C16:0 platelet activating factor, a lipid that accumulates intraneuronally in Alzheimer's patients
RFX1	4089	-1.353	-0.551	Major transcriptional repressor of DNA-damage-regulated genes; recruits repressors Tup1p and Cyc8p to their promoters; involved in DNA damage and replication checkpoint pathway; similar to a family of mammalian DNA binding RFX1-4 proteins
SAT4	4088	-1.352	-0.871	Ser/Thr protein kinase involved in salt tolerance; funtions in regulation of Trk1p-Trk2p potassium transporter; overexpression affects the Fe-S and lipoamide containing proteins in the mitochondrion; required for lipoylation of Lat1p, Kgd2p and Gcv3p; partially redundant with Hal5p; has similarity to Npr1p; localizes to the cytoplasm and mitochondrion
SWI6	4086	-1.349	-2.301	Transcription cofactor; forms complexes with Swi4p and Mbp1p to regulate transcription at the G1/S transition; involved in meiotic gene expression; also binds Stb1p to regulate transcription at START; cell wall stress induces phosphorylation by Mpk1p, which regulates Swi6p localization; required for the unfolded protein response, independently of its known transcriptional coactivators
LAG1	4087	-1.349	0.883	Ceramide synthase component; involved in synthesis of ceramide from C26(acyl)-coenzyme A and dihydrosphingosine or phytosphingosine, functionally equivalent to Lac1p; forms ER foci upon DNA replication stress; homolog of human CERS2, a tumor metastasis suppressor gene whose silencing enhances invasion/metastasis of prostate cancer cells; LAG1 has a paralog, LAC1, that arose from the whole genome duplication
RPA49	4085	-1.343	-1.994	RNA polymerase I subunit A49; essential for nucleolar assembly and for high polymerase loading rate; required for nucleolar localization of Rpa34p
APS2	4084	-1.343	0.006	Small subunit of the clathrin-associated adaptor complex AP-2; AP-2 is involved in protein sorting at the plasma membrane; related to the sigma subunit of the mammalian plasma membrane clathrin-associated protein (AP-2) complex
STP22	4083	-1.342	0.209	Component of the ESCRT-I complex; ESCRT-I is involved in ubiquitin-dependent sorting of proteins into the endosome; prevents polyubiquitination of the arrestin-related protein Rim8p, thereby directing its monoubiquitination by Rsp5p; homologous to the mouse and human Tsg101 tumor susceptibility gene; mutants exhibit a Class E Vps phenotype
DHH1	4082	-1.341	-0.868	Cytoplasmic DEAD-box helicase, stimulates mRNA decapping; coordinates distinct steps in mRNA function and decay, interacting with both decapping and deadenylase complexes; role in translational repression, mRNA decay, and possibly mRNA export; interacts and cooperates with Ngr1p to promote specific mRNA decay; ATP- and RNA-bound form promotes processing body (PB)

				assembly, while ATPase stimulation by Not1p promotes PB disassembly; forms cytoplasmic foci on replication stress
SNF4	4081	-1.34	-1.21	Activating gamma subunit of the AMP-activated Snf1p kinase complex; additional subunits of the complex are Snf1p and a Sip1p/Sip2p/Gal83p family member; activates glucose-repressed genes, represses glucose-induced genes; role in sporulation, and peroxisome biogenesis; protein abundance increases in response to DNA replication stress
LST4	4080	-1.338	-1.287	Subunit of the Lst4p-Lst7p GTPase activating protein complex for Gtr2p; stimulates the GTPase activity of Rag family GTPase Gtr2p, within the context of the Gtr1p-Gtr2p heterodimer, after amino acid stimulation; required for activation of TORC1 in response to amino acid stimulation; recruited to the vacuolar membrane during amino acid starvation and released from the membrane by TORC1; required for the transport of amino acid permease Gap1p from the Golgi to the cell surface
RPL21B	4079	-1.332	-0.512	Ribosomal 60S subunit protein L21B; homologous to mammalian ribosomal protein L21, no bacterial homolog; RPL21B has a paralog, RPL21A, that arose from the whole genome duplication
SER1	4078	-1.33	-0.841	3-phosphoserine aminotransferase; catalyzes the formation of phosphoserine from 3-phosphohydroxypyruvate, required for serine and glycine biosynthesis; regulated by the general control of amino acid biosynthesis mediated by Gcn4p; protein abundance increases in response to DNA replication stress
VMA8	4077	-1.324	-1.574	Subunit D of the V1 peripheral membrane domain of V-ATPase; part of the electrogenic proton pump found throughout the endomembrane system; plays a role in the coupling of proton transport and ATP hydrolysis; the V1 peripheral membrane domain of the vacuolar H+-ATPase (V-ATPase) has eight subunits
VPS45	4076	-1.314	-1.171	Protein of the Sec1p/Munc-18 family; essential for vacuolar protein sorting; required for the function of Pep12p and the early endosome/late Golgi SNARE Tlg2p; essential for fusion of Golgi-derived vesicles with the prevacuolar compartment
IKS1	4074	-1.307	-0.41	Protein kinase of unknown cellular role; putative serine/threonine kinase; expression is induced during mild heat stress; deletion mutants are hypersensitive to copper sulphate and resistant to sorbate; interacts with an N-terminal fragment of Sst2p
HIS5	4075	-1.307	0.035	Histidinol-phosphate aminotransferase; catalyzes the seventh step in histidine biosynthesis; responsive to general control of amino acid biosynthesis; mutations cause histidine auxotrophy and sensitivity to Cu, Co, and Ni salts
RTT109	4072	-1.305	-1.422	Histone acetyltransferase; critical for cell survival in presence of DNA damage during S phase, required for recovery after DSB repair; acetylates H3K56, H3K9; H3K56 acetylation activity required for expression homeostasis, buffering of mRNA synthesis rate against changes in gene dosage during S phase; involved in non-homologous end joining and regulation of Ty1 transposition; prevents hyper-amplification of rDNA; interacts physically with Vps75p
YGR042W	4073	-1.305	0.087	hypothetical protein; involved in maintenance of proper telomere length; green fluorescent protein (GFP)-fusion protein localizes to both the cytoplasm and the nucleus; forms nuclear foci upon DNA replication stress
SSF1	4071	-1.303	-0.818	Constituent of 66S pre-ribosomal particles; required for ribosomal large subunit maturation; functionally redundant with Ssf2p; member of the Brix family; SSF1 has a paralog, SSF2, that arose from the whole genome duplication
RPL6B	4070	-1.3	-1.811	Ribosomal 60S subunit protein L6B; binds 5.8S rRNA; homologous to mammalian ribosomal protein L6, no bacterial homolog; RPL6B has a paralog, RPL6A, that arose from the whole genome duplication
LYS1	4069	-1.299	-1.462	Saccharopine dehydrogenase (NAD+, L-lysine-forming); catalyzes the conversion of saccharopine to L-lysine, which is the final step in the lysine biosynthesis pathway; also has mRNA binding activity
SAC1	4068	-1.292	-1.83	Phosphatidylinositol phosphate (PtdInsP) phosphatase; involved in hydrolysis of PtdIns[4]P in the early and medial Golgi; regulated by interaction with Vps74p; ER localized transmembrane protein which cycles through the Golgi; involved in protein trafficking and processing, secretion, and cell wall maintenance; regulates sphingolipid biosynthesis through the modulation of PtdIns(4)P metabolism

RPS24A	4066	-1.291	-1.058	Protein component of the small (40S) ribosomal subunit; homologous to mammalian ribosomal protein S24, no bacterial homolog; RPS24A has a paralog, RPS24B, that arose from the whole genome duplication
SEC66	4067	-1.291	-1.949	Non-essential subunit of Sec63 complex; with Sec61 complex, Kar2p/BiP and Lhs1p forms a channel competent for SRP-dependent and post-translational SRP-independent protein targeting and import into the ER; other members are Sec63p, Sec62p, and Sec72p
GLR1	4065	-1.288	0.573	Cytosolic and mitochondrial glutathione oxidoreductase; converts oxidized glutathione to reduced glutathione; cytosolic Glr1p is the main determinant of the glutathione redox state of the mitochondrial intermembrane space; mitochondrial Glr1p has a role in resistance to hyperoxia; protein abundance increases in response to DNA replication stress
TPS1	4064	-1.287	-1.499	Synthase subunit of trehalose-6-P synthase/phosphatase complex; synthesizes the storage carbohydrate trehalose, which is critically important for survival of long-term desiccation; also found in a monomeric form; expression is induced by the stress response and repressed by the Ras-cAMP pathway; protein abundance increases in response to DNA replication stress and in response to prolonged exposure to boric acid
LGE1	4063	-1.285	-0.867	hypothetical protein; null mutant forms abnormally large cells, and homozygous diploid null mutant displays delayed premeiotic DNA synthesis and reduced efficiency of meiotic nuclear division
SLM6	4061	-1.278	-0.581	No Description Found
VPS36	4062	-1.278	-0.587	Component of the ESCRT-II complex; contains the GLUE (GRAM Like Ubiquitin binding in EAP45) domain which is involved in interactions with ESCRT-I and ubiquitin-dependent sorting of proteins into the endosome; plays a role in the formation of mutant huntingtin (Htt) aggregates in yeast
HFI1	4060	-1.277	-2.639	Adaptor protein required for structural integrity of the SAGA complex; a histone acetyltransferase-coactivator complex that is involved in global regulation of gene expression through acetylation and transcription functions
SRN2	4059	-1.273	0.061	Component of the ESCRT-I complex; ESCRT-I is involved in ubiquitin-dependent sorting of proteins into the endosome; suppressor of rna1-1 mutation; may be involved in RNA export from nucleus
RPS27A	4058	-1.272	0.176	Protein component of the small (40S) ribosomal subunit; homologous to mammalian ribosomal protein S27, no bacterial homolog; RPS27A has a paralog, RPS27B, that arose from the whole genome duplication; protein abundance increases in response to DNA replication stress
THO2	4057	-1.269	-0.496	Subunit of the THO complex; THO is required for efficient transcription elongation and involved in transcriptional elongation-associated recombination; required for LacZ RNA expression from certain plasmids
NUP2	4056	-1.269	1.139	Nucleoporin involved in nucleocytoplasmic transport; binds to either the nucleoplasmic or cytoplasmic faces of the nuclear pore complex depending on Ran-GTP levels; also has a role in chromatin organization
SER2	4053	-1.267	-1.344	Phosphoserine phosphatase of the phosphoglycerate pathway; involved in serine and glycine biosynthesis, expression is regulated by the available nitrogen source
DLD1	4055	-1.267	-0.07	Major mitochondrial D-lactate dehydrogenase; oxidizes D-lactate to pyruvate, transcription is heme-dependent, repressed by glucose, and derepressed in ethanol or lactate; located in the mitochondrial inner membrane
YSY6	4054	-1.267	-0.607	hypothetical protein; expression suppresses a secretory pathway mutation in E. coli; has similarity to the mammalian RAMP4 protein involved in secretion
URM1	4052	-1.263	-1.341	Ubiquitin-like protein involved in thiolation of cytoplasmic tRNAs; receives sulfur from the E1-like enzyme Uba4p and transfers it to tRNA; also functions as a protein tag with roles in nutrient sensing and oxidative stress response
RSM7	4051	-1.26	-1.879	Mitochondrial ribosomal protein of the small subunit; has similarity to E. coli S7 ribosomal protein
CGI121	4050	-1.258	-0.822	Component of the EKC/KEOPS complex; EKC/KEOPS complex is required for t6A tRNA modification and telomeric TG1-3 recombination; may have role in transcription; Cgi121p is dispensable for tRNA modification; other complex members are Bud32p, Kae1p, Pcc1p, and Gon7p

FLX1	4049	-1.253	-0.235	Mitochondrial flavin adenine dinucleotide transporter; FAD is a synthesis product of riboflavin; human homolog SLC25A32 is implicated in multiple acyl-CoA dehydrogenase deficiency (MADD) or glutaric aciduria type II (GAII), and can complement yeast null mutant
OCA1	4048	-1.249	-0.728	Putative protein tyrosine phosphatase; required for cell cycle arrest in response to oxidative damage of DNA
DOC1	4047	-1.244	-1.207	Processivity factor; required for the ubiquitination activity of the anaphase promoting complex (APC), mediates the activity of the APC by contributing to substrate recognition; involved in cyclin proteolysis; contains a conserved DOC1 homology domain
DID2	4046	-1.243	-0.149	Class E protein of the vacuolar protein-sorting (Vps) pathway; binds Vps4p and directs it to dissociate ESCRT-III complexes; forms a functional and physical complex with Ist1p; human ortholog may be altered in breast tumors
GLN3	4045	-1.243	-0.667	Transcriptional activator of genes regulated by nitrogen catabolite repression; localization and activity regulated by quality of nitrogen source and Ure2p
TRM7	4044	-1.241	-0.301	2'-O-ribose methyltransferase; methylates the 2'-O-ribose of tRNA-Phe, tRNA-Trp, and tRNA-Leu at positions C32 and N34 of tRNA anticodon loop; crucial biological role likely modification of tRNA-Phe; interacts with Trm732p and Rtt10p in 2'-O-methylation of C32 and N34 substrate tRNAs, respectively; yeast null mutant can be functionally complemented by human FTSJ1, mutations in which have been implicated in nonsyndromic X-linked intellectual disability (NSXLID)
ARG3	4043	-1.238	-1.756	Ornithine carbamoyltransferase; also known as carbamoylphosphate:L-ornithine carbamoyltransferase; catalyzes the biosynthesis of the arginine precursor citrulline
MMP1	4042	-1.234	-0.389	High-affinity S-methylmethionine permease; required for utilization of S-methylmethionine as a sulfur source; has similarity to S-adenosylmethionine permease Sam3p
CBC2	4040	-1.233	-1.296	Small subunit of the heterodimeric cap binding complex with Sto1p; interacts with Npl3p, possibly to package mRNA for export from the nucleus; may have a role in telomere maintenance; contains an RNA-binding motif
RPS0B	4041	-1.233	-0.988	Protein component of the small (40S) ribosomal subunit; RPS0B has a paralog, RPS0A, that arose from the whole genome duplication; required for maturation of 18S rRNA along with Rps0Ap; deletion of either RPS0 gene reduces growth rate, deletion of both genes is lethal; homologous to human ribosomal protein SA and bacterial S2
ALF1	4039	-1.232	-1.19	Alpha-tubulin folding protein; similar to mammalian cofactor B; Alf1p-GFP localizes to cytoplasmic microtubules; required for the folding of alpha-tubulin and may play an additional role in microtubule maintenance
4.00 SIT	4038	-1.231	-2.428	Ceramide-activated, type 2A-related serine-threonine phosphatase; functions in G1/S transition of mitotic cycle; controls lifespan, mitochondrial function, cell cycle progression by regulating HXK2 phosphorylation; regulator of COPII coat dephosphorylation; required for ER to Golgi traffic; interacts with Hrr25p kinase; cytoplasmic and nuclear protein that modulates functions mediated by Pkc1p including cell wall and actin cytoskeleton organization; similar to human PP6
MET5	4037	-1.231	1.113	Sulfite reductase beta subunit; involved in amino acid biosynthesis, transcription repressed by methionine
RPS23A	4034	-1.23	-0.885	Ribosomal protein 28 (rp28) of the small (40S) ribosomal subunit; required for translational accuracy; homologous to mammalian ribosomal protein S23 and bacterial S12; RPS23A has a paralog, RPS23B, that arose from the whole genome duplication; deletion of both RPS23A and RPS23B is lethal
BUD20	4036	-1.23	-1.928	C2H2-type zinc finger protein required for ribosome assembly; shuttling factor which associates with pre-60S particles in the nucleus, accompanying them to the cytoplasm; cytoplasmic dissociation of Bud20p requires Drg1p; N-terminus harbors a nuclear localization signal (NLS) and a nuclear export signal (NES); cytoplasmic Bud20p is reimported by Kap123-dependent pathway; involved in bud-site selection; diploid mutants display a random budding pattern; similar to human ZNF593
HIS6	4035	-1.23	-0.124	Enzyme that catalyzes the fourth step in the histidine pathway; Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase; mutations cause histidine auxotrophy and sensitivity to Cu, Co, and Ni salts

PAR32	4033	-1.226	-1.285	hypothetical protein; hyperphosphorylated upon rapamycin treatment in a Tap42p-dependent manner; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm; PAR32 is not an essential gene
RAD52	4031	-1.218	-1.035	Protein that stimulates strand exchange; stimulates strand exchange by facilitating Rad51p binding to single-stranded DNA; anneals complementary single-stranded DNA; involved in the repair of double-strand breaks in DNA during vegetative growth and meiosis and UV induced sister chromatid recombination
YCL001W-A	4032	-1.218	-0.19	hypothetical protein; YCL001W-A gene has similarity to DOM34 and is present in a region duplicated between chromosomes XIV and III
GLY1	4030	-1.209	-1.637	Threonine aldolase; catalyzes the cleavage of L-allo-threonine and L-threonine to glycine; involved in glycine biosynthesis
YJL070C	4029	-1.205	-0.295	Putative metallo-dependent hydrolase superfamily protein; similar to AMP deaminases but lacks key catalytic residues and does not rescue purine nucleotide metabolic defect of quadruple aah1 ade8 amd1 his1 mutant; may regulate purine nucleotide homeostasis as overexpression in an AMD1 strain grown in adenine results in greatly reduced GDP and GTP intracellular levels; not an essential gene; YJL070C has a paralog, YBR284W, that arose from the whole genome duplication
RPL20A	4027	-1.203	-1.543	Ribosomal 60S subunit protein L20A; homologous to mammalian ribosomal protein L18A, no bacterial homolog; RPL20A has a paralog, RPL20B, that arose from the whole genome duplication
WSS1	4028	-1.203	-0.451	SUMO-ligase and SUMO-targeted metalloprotease; involved in DNA repair; removes DNA-protein crosslinks at stalled replication forks during replication of damaged DNA; clears chromatin-bound sumoylated proteins; localizes to single spot on nuclear periphery of mother cells but not daughters; exhibits vacuolar localization upon genotoxic stress; activated by DNA binding; member of minigluzincins protease family with mammalian DVC1/Spartan

SUPPLEMENTARY TABLE S7: Failing Genes in Betaxanthin Screen

gene	Yield_Rank	Yield_Score	Size_Score	Description
YOR186W	na	na	na	hypothetical protein; proper regulation of expression during heat stress is sphingolipid-dependent; mCherry fusion protein localizes to the vacuole; YOR186W has a paralog, YLR297W, that arose from the whole genome duplication
STE11	na	na	na	Signal transducing MEK kinase; involved in pheromone response and pseudohyphal/invasive growth pathways where it phosphorylates Ste7p, and the high osmolarity response pathway, via phosphorylation of Pbs2p; regulated by Ste20p and Ste50p; protein abundance increases in response to DNA replication stress
YHR177W	na	na	na	Putative transcription factor containing a WOPR domain; binds DNA in vitro; similar to C. albicans Wor1p transcription factor that regulates white-opaque switching; overexpression causes a cell cycle delay or arrest
FUS3	na	na	na	Mitogen-activated serine/threonine protein kinase involved in mating; phosphoactivated by Ste7p; substrates include Ste12p, Far1p, Bni1p, Sst2p; inhibits invasive growth during mating by phosphorylating Tec1p, promoting its; inhibits recruitment of Ste5p, Cdc42p-mediated asymmetry and mating morphogenesis
KAR4	na	na	na	Transcription factor required for response to pheromones; also required during meiosis; exists in two forms, a slower-migrating form more abundant during vegetative growth and a faster-migrating form induced by pheromone
RIB1	na	na	na	GTP cyclohydrolase II; catalyzes the first step of the riboflavin biosynthesis pathway
DSE2	na	na	na	Daughter cell-specific secreted protein with similarity to glucanases; degrades cell wall from the daughter side causing daughter to separate from mother; expression is repressed by cAMP
IMP2'	na	na	na	Transcriptional activator involved in maintenance of ion homeostasis; also involved in protection against DNA damage caused by bleomycin and other oxidants; contains a C-terminal leucine-rich repeat

CDC26	na	na	na	Subunit of the Anaphase-Promoting Complex/Cyclosome (APC/C); which is a ubiquitin-protein ligase required for degradation of anaphase inhibitors, including mitotic cyclins, during the metaphase/anaphase transition; relocalizes to the cytosol in response to hypoxia
TIR3	na	na	na	Cell wall mannoprotein; member of Srp1p/Tip1p family of serine-alanine-rich proteins; expressed under anaerobic conditions and required for anaerobic growth; TIR3 has a paralog, TIR2, that arose from the whole genome duplication
MGT1	na	na	na	DNA repair methyltransferase (6-O-methylguanine-DNA methylase); involved in protection against DNA alkylation damage
YOR105W	na	na	na	hypothetical protein; expressed at both mRNA and protein levels
FYV7	na	na	na	Essential protein required for maturation of 18S rRNA; required for survival upon exposure to K1 killer toxin
CEX1	na	na	na	Component of nuclear aminoacylation-dependent tRNA export pathway; cytoplasmic; interacts with nuclear pore component Nup116p; copurifies with tRNA export receptors Los1p and Msn5p, as well as eIF-1a; required for activation of RAN GTPase Gsp1p and dissociation of receptor-tRNA-Gsp1p export complex; recruits Rna1p from cytoplasm to NPC, facilitates Rna1p activation of Gsp1p GTPase activity by enabling Rna1p to gain access to Gsp1p-GTP bound to export receptor tRNA complex
YFL035C-B	na	na	na	No Description Found
RPS24B	na	na	na	Protein component of the small (40S) ribosomal subunit; homologous to mammalian ribosomal protein S24, no bacterial homolog; RPS24B has a paralog, RPS24A, that arose from the whole genome duplication
YOR314W	na	na	na	hypothetical protein; conserved across S. cerevisiae strains
MNI1	na	na	na	No Description Found
MFA1	na	na	na	Mating pheromone a-factor; made by a cells; interacts with alpha cells to induce cell cycle arrest and other responses leading to mating; biogenesis involves C-terminal modification, N-terminal proteolysis, and export; also encoded by MFA2
STE20	na	na	na	Cdc42p-activated signal transducing kinase; involved in pheromone response, pseudohyphal/invasive growth, vacuole inheritance, down-regulation of sterol uptake; GBB motif binds Ste4p; member of the PAK (p21-activated kinase) family
CLC1	na	na	na	Clathrin light chain; subunit of the major coat protein involved in intracellular protein transport and endocytosis; regulates endocytic progression; thought to regulate clathrin function; the clathrin triskelion is a trimeric molecule composed of three heavy chains that radiate from a vertex and three light chains which bind noncovalently near the vertex of the triskelion
MBB1	na	na	na	hypothetical protein; conserved among S. cerevisiae strains, not conserved in closely related Saccharomyces species; protein detected in large-scale protein-protein interaction studies; YJL199C is not an essential gene
HTL1	na	na	na	Component of the RSC chromatin remodeling complex; RSC functions in transcriptional regulation and elongation, chromosome stability, and establishing sister chromatid cohesion; involved in telomere maintenance
ERG4	na	na	na	C-24(28) sterol reductase; catalyzes the final step in ergosterol biosynthesis; mutants are viable, but lack ergosterol
DIG1	na	na	na	MAP kinase-responsive inhibitor of the Ste12p transcription factor; involved in the regulation of mating-specific genes and the invasive growth pathway; related regulators Dig1p and Dig2p bind to Ste12p; DIG1 has a paralog, DIG2, that arose from the whole genome duplication
TUM1	na	na	na	Rhodanese domain sulfur transferase; accepts persulfite from Nfs1p and transfers it to Uba4p in the pathway for 2-thiolation of the wobble uridine base of tRNAs; also stimulates sulfur transfer by Nfs1p; may be mitochondrially localized
PEP7	na	na	na	Adaptor protein involved in vesicle-mediated vacuolar protein sorting; multivalent adaptor protein; facilitates vesicle-mediated vacuolar protein sorting by ensuring high-fidelity vesicle docking and fusion, which are essential for targeting of vesicles to the endosome; required for vacuole inheritance

QDR2	na	na	na	Plasma membrane transporter of the major facilitator superfamily; member of the 12-spanner drug:H(+) antiporter DHA1 family; exports copper; has broad substrate specificity and can transport many mono- and divalent cations; transports a variety of drugs and is required for resistance to quinidine, barban, cisplatin, and bleomycin; contributes to potassium homeostasis; expression is regulated by copper
FUN12	na	na	na	Translation initiation factor eIF5B; GTPase that promotes Met-tRNAiMet binding to ribosomes and ribosomal subunit joining; promotes GTP-dependent maturation of 18S rRNA by Nob1p; protein abundance increases in response to DNA replication stress; homolog of bacterial IF2
SIR4	na	na	na	SIR protein involved in assembly of silent chromatin domains; silent information regulator (SIR) along with SIR2 and SIR3; involved in assembly of silent chromatin domains at telomeres and the silent mating-type loci; some alleles of SIR4 prolong lifespan; required for telomere hypercluster formation in quiescent yeast cells
PAF1	na	na	na	Component of the Paf1p complex involved in transcription elongation; binds to and modulates the activity of RNA polymerases I and II; required for expression of a subset of genes, including cell cycle-regulated genes; involved in SER3 repression by helping to maintain SRG1 transcription-dependent nucleosome occupancy; homolog of human PD2/hPAF1
CST26	na	na	na	Acyltransferase; enzyme mainly responsible for the introduction of saturated very long chain fatty acids into neo-synthesized molecules of phosphatidylinositol; required for incorporation of stearic acid into phosphatidylinositol; affects chromosome stability when overexpressed; CST26 has a paralog, YDR018C, that arose from the whole genome duplication
AXL1	na	na	na	Haploid specific endoprotease of a-factor mating pheromone; performs one of two N-terminal cleavages during maturation of a-factor mating pheromone; required for axial budding pattern of haploid cells
NCR1	na	na	na	Vacuolar membrane protein; transits through the biosynthetic vacuolar protein sorting pathway, involved in sphingolipid metabolism; cells lacking Ncr1p exhibit high levels of long chain bases (LCB), similar to the accumulation of high amounts of lipids observed in patients with Neimann-Pick C, a disease caused by loss-of-function mutations in NPC1, the functional ortholog of Ncr1p
SHS1	na	na	na	Component of the septin ring that is required for cytokinesis; present at the ends of rod-like septin hetero-oligomers; C-terminal extension is important for recruitment of Bni5p to the mother-bud neck, which in turn is required for Myo1p recruitment and cytokinesis; undergoes sumoylation and phosphorylation during mitosis; protein abundance increases in response to DNA replication stress
CYC8	na	na	na	General transcriptional co-repressor; acts together with Tup1p; also acts as part of a transcriptional co-activator complex that recruits the SWI/SNF and SAGA complexes to promoters; can form the prion [OCT+]
REV1	na	na	na	Deoxycytidyl transferase; involved in repair of abasic sites and adducted guanines in damaged DNA by translesion synthesis (TLS); forms a complex with the subunits of DNA polymerase zeta, Rev3p and Rev7p; relocalizes from nucleus to cytoplasm upon DNA replication stress
VPS38	na	na	na	Part of a Vps34p phosphatidylinositol 3-kinase complex; functions in carboxypeptidase Y (CPY) sorting; binds Vps30p and Vps34p to promote production of phosphatidylinositol 3-phosphate (PtdIns3P) which stimulates kinase activity; required for overflow degradation of misfolded proteins when ERAD is saturated
SPT8	na	na	na	Subunit of the SAGA transcriptional regulatory complex; not present in SAGA-like complex SLIK/SALSA; required for SAGA-mediated inhibition at some promoters
APL2	na	na	na	Beta-adaptin subunit of the clathrin-associated protein (AP-1) complex; binds clathrin; involved in clathrin-dependent Golgi protein sorting; protein abundance increases in response to DNA replication stress
RAD2	na	na	na	Single-stranded DNA endonuclease; cleaves single-stranded DNA during nucleotide excision repair to excise damaged DNA; subunit of Nucleotide Excision Repair Factor 3 (NEF3); homolog of human XPG protein
YNR068C	na	na	na	hypothetical protein; exhibits homology to C-terminal end of Bul1p; expressed as a readthrough product of BSC5, the readthrough locus being termed BUL3; the BUL3 readthrough product is involved in ubiquitin-mediated sorting of plasma

				membrane proteins and interacts with WW domains of Rsp5p in vitro, but in a functionally different way than the non-readthrough form
ECM31	na	na	na	Ketopantoate hydroxymethyltransferase; required for pantothenic acid biosynthesis, converts 2-oxoisovalerate into 2-dehydropantoate
STE2	na	na	na	Receptor for alpha-factor pheromone; seven transmembrane-domain GPCR that interacts with both pheromone and a heterotrimeric G protein to initiate the signaling response that leads to mating between haploid a and alpha cells
CLB1	na	na	na	B-type cyclin involved in cell cycle progression; activates Cdc28p to promote the transition from G2 to M phase; accumulates during G2 and M, then targeted via a destruction box motif for ubiquitin-mediated degradation by the proteasome; CLB1 has a paralog, CLB2, that arose from the whole genome duplication
KEX1	na	na	na	Cell death protease essential for hypochlorite-induced apoptosis; involved in the processing of killer toxin and alpha factor precursor; cleaves Lys and Arg residues from the C-terminus of peptides and proteins
UBX6	na	na	na	UBX (ubiquitin regulatory X) domain-containing protein; interacts with Cdc48p, transcription is repressed when cells are grown in media containing inositol and choline; UBX6 has a paralog, UBX7, that arose from the whole genome duplication
VPS33	na	na	na	ATP-binding protein that is a subunit of the HOPS and CORVET complexes; essential for protein sorting, vesicle docking, and fusion at the vacuole; binds to SNARE domains
APQ13	na	na	na	No Description Found
NPL6	na	na	na	Component of the RSC chromatin remodeling complex; interacts with Rsc3p, Rsc30p, Ldb7p, and Htl1p to form a module important for a broad range of RSC functions
BAR1	na	na	na	Aspartyl protease; secreted into the periplasmic space of mating type a cell; helps cells find mating partners; cleaves and inactivates alpha factor allowing cells to recover from alpha-factor-induced cell cycle arrest
ATG11	na	na	na	Adapter protein for pexophagy and the Cvt targeting pathway; directs receptor-bound cargo to the phagophore assembly site (PAS) for packaging into vesicles; required for recruiting other proteins to the PAS; recruits Dnm1p to facilitate fission of mitochondria that are destined for removal by mitophagy
BUD16	na	na	na	Putative pyridoxal kinase; a key enzyme involved in pyridoxal 5'-phosphate synthesis, the active form of vitamin B6; required for genome integrity; involved in bud-site selection; similarity to yeast BUD17 and human pyridoxal kinase (PDXK)
INH1	na	na	na	Protein that inhibits ATP hydrolysis by the F1F0-ATP synthase; inhibitory function is enhanced by stabilizing proteins Stf1p and Stf2p; has a calmodulin-binding motif and binds calmodulin in vitro; INH1 has a paralog, STF1, that arose from the whole genome duplication
YKL069W	na	na	na	Methionine-R-sulfoxide reductase; reduces the R enantiomer of free Met-SO, in contrast to Ycl033Cp which reduces Met-R-SO in a peptide linkage; has a role in protection against oxidative stress; relative distribution to the nucleus increases upon DNA replication stress
ISA2	na	na	na	Protein required for maturation of mitochondrial [4Fe-4S] proteins; functions in a complex with Isa1p and possibly Iba57p; localizes to the mitochondrial intermembrane space, overexpression of ISA2 suppresses grx5 mutations
COX9	na	na	na	Subunit VIIa of cytochrome c oxidase (Complex IV); Complex IV is the terminal member of the mitochondrial inner membrane electron transport chain
UGA1	na	na	na	Gamma-aminobutyrate (GABA) transaminase; also known as 4-aminobutyrate aminotransferase; involved in the 4-aminobutyrate and glutamate degradation pathways; required for normal oxidative stress tolerance and nitrogen utilization; protein abundance increases in response to DNA replication stress
MIP6	na	na	na	Putative RNA-binding protein; interacts with Mex67p, which is a component of the nuclear pore involved in nuclear mRNA export; MIP6 has a paralog, PES4, that arose from the whole genome duplication
YER158C	na	na	na	hypothetical protein; potentially phosphorylated by Cdc28p; YER158C has a paralog, AFR1, that arose from the whole genome duplication

Chapter 7

Appendix II: Supplementary Material for: *Microbial cell factory optimisation using genome-wide host-pathway interaction screens*

Supplementary materials

Microbial cell factory optimisation using genome-wide host-pathway interaction screens

Paul Cachera[1], Nikolaj Can Kurt[2], Andreas Røpke[2], Tomas Strucko[2], Uffe H. Mortensen[2], Michael K. Jensen[1,#]

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark

[2] Department of Biotechnology and Biomedicine, Technical University of Denmark, Kgs. Lyngby, Denmark

[#] = to whom correspondence should be addressed: Michael K. Jensen mije@biosustain.dtu.dk

Supplementary figure S1 Cloning of the CCM cassette.

Supplementary figure S2: Biological replicate of experiments in figure 2.

Supplementary figure S3: Raw, uncorrected and corrected CRI-SPA data.

Supplementary figure S4: STRING Clusters for 250 most extreme gene groups.

Supplementary figure S5: K-clique clusters extracted from STRING networks and annotated with GOEA.

Supplementary Figure S6: CRI-SPA fluorescence score correlates with CCM titers in liquid culture.

Supplementary Figure S7: Biological replicate of experiment in figure 5.

Supplementary Table S1 Sourcing of parts used to build the CCM cassette.

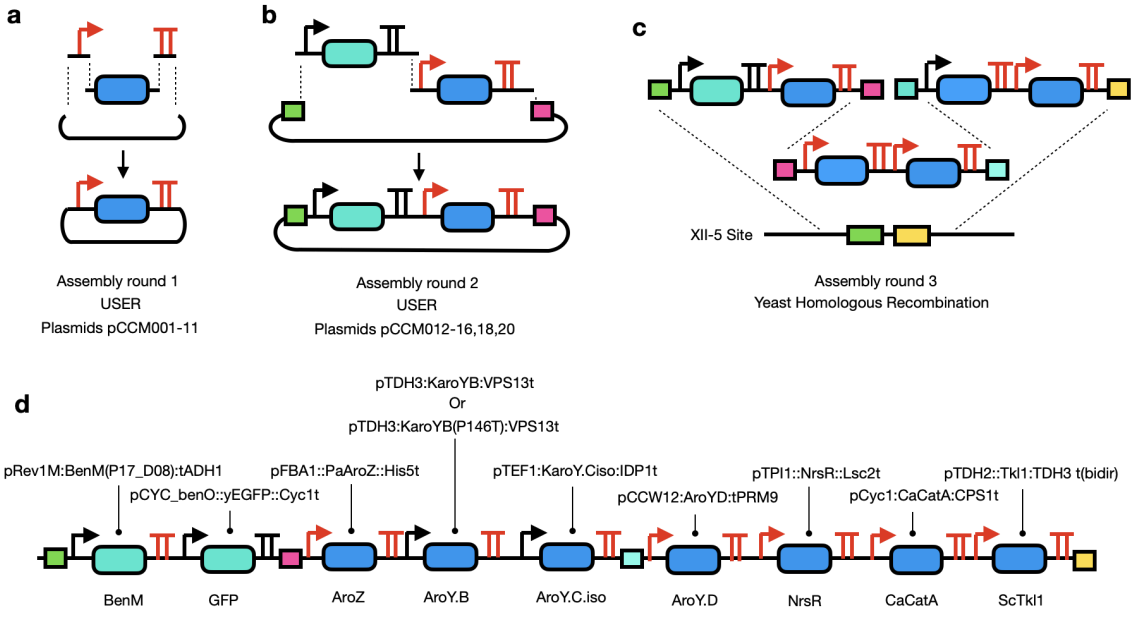
Supplementary Table S2 Gene fluorescent ranking for the KO and OEx Screens

Supplementary Table S3 Gene hits selected for further testing.

Supplementary Table S4 Strains used in this study

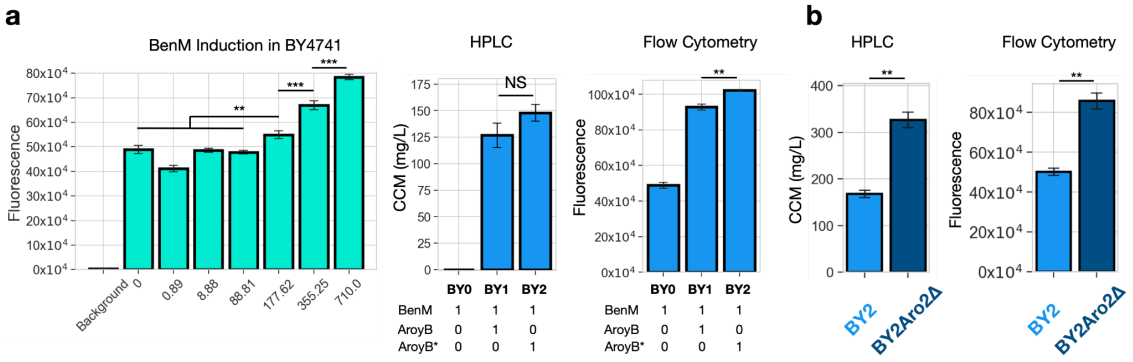
Supplementary Table S5 Plasmids used in this study

Supplementary Table S6 Primers used in this study

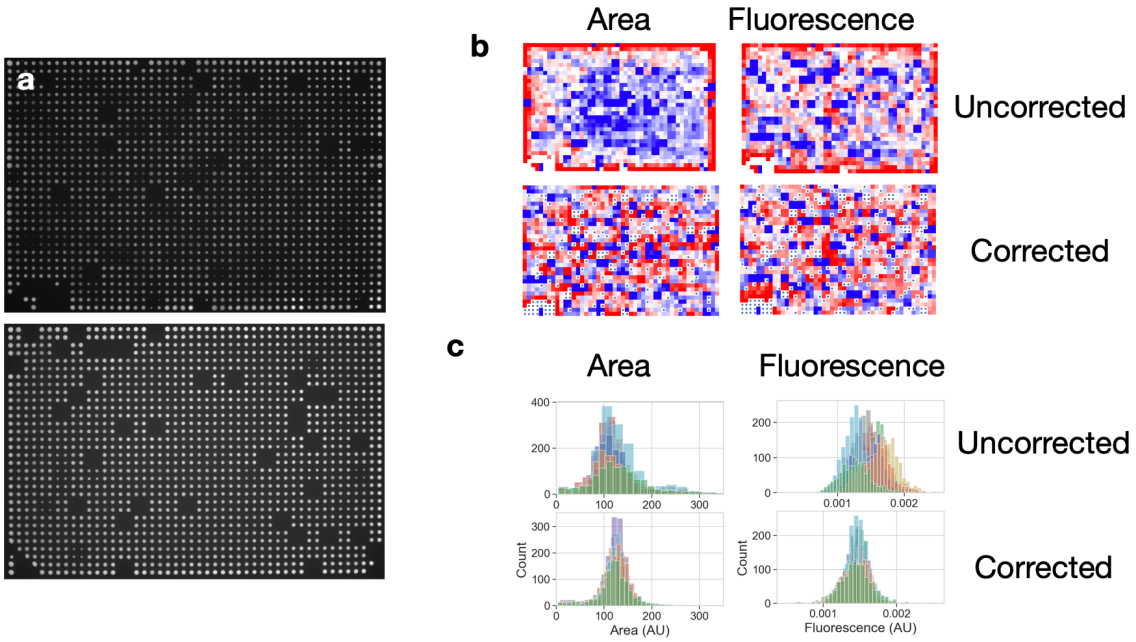


Supplementary Figure S1. Refactoring the CCM cassette. a) The CCM cassette in which repetitive promoters and terminators were replaced was built in three rounds of assembly. **a)** Individual parts (promoters, CDSs, terminators) were PCR amplified and USER-assembled into plasmids pCCM001-11. **b)** Expression cassettes from plasmids pCCM001-11 were PCR

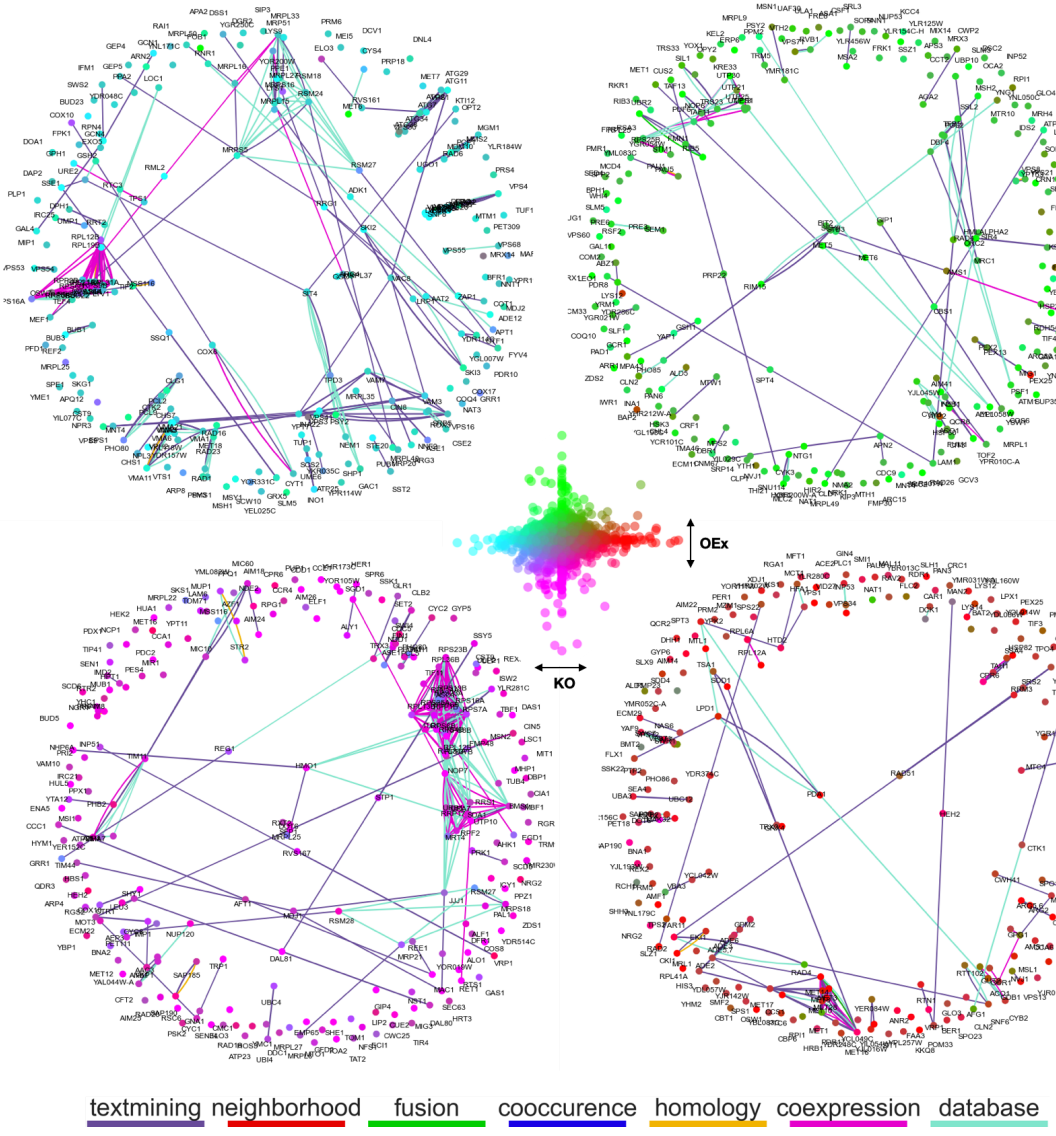
amplified and assembled into joined expression cassettes flanked by matching homology sequences for assembly in yeast. This yielded plasmids pCCM012-16,18, and 20. **c)** The joined expression cassettes from plasmids pCCM012-16,18,20 were excised and assembled by homologous recombination at integration site XII-5 in yeast. **d)** All parts in the final CCM expression cassette. The replaced promoters and terminators are marked in red. All parts and their sources are listed in **Supplementary Table S2**. All plasmids and primers used are listed in **Supplementary Tables S4-S5,** respectively.



Supplementary Figure S2: Biological replicate of experiments in Figure 2. a) Left: induction of the biosensor in strain BY0. Centre and right: CCM titers and fluorescence measured by flow cytometry for BY0, BY1 and BY2. **b**) KO of positive control gene *Aro4* results in significant increase in CCM titers. The increase in CCM titers caused by *Aro2* causes a fluorescent change which can be detected by flow cytometry. For all graphs, significance was assessed with a two-sided t-test *<0.05, **<0.01, ***<0.001.

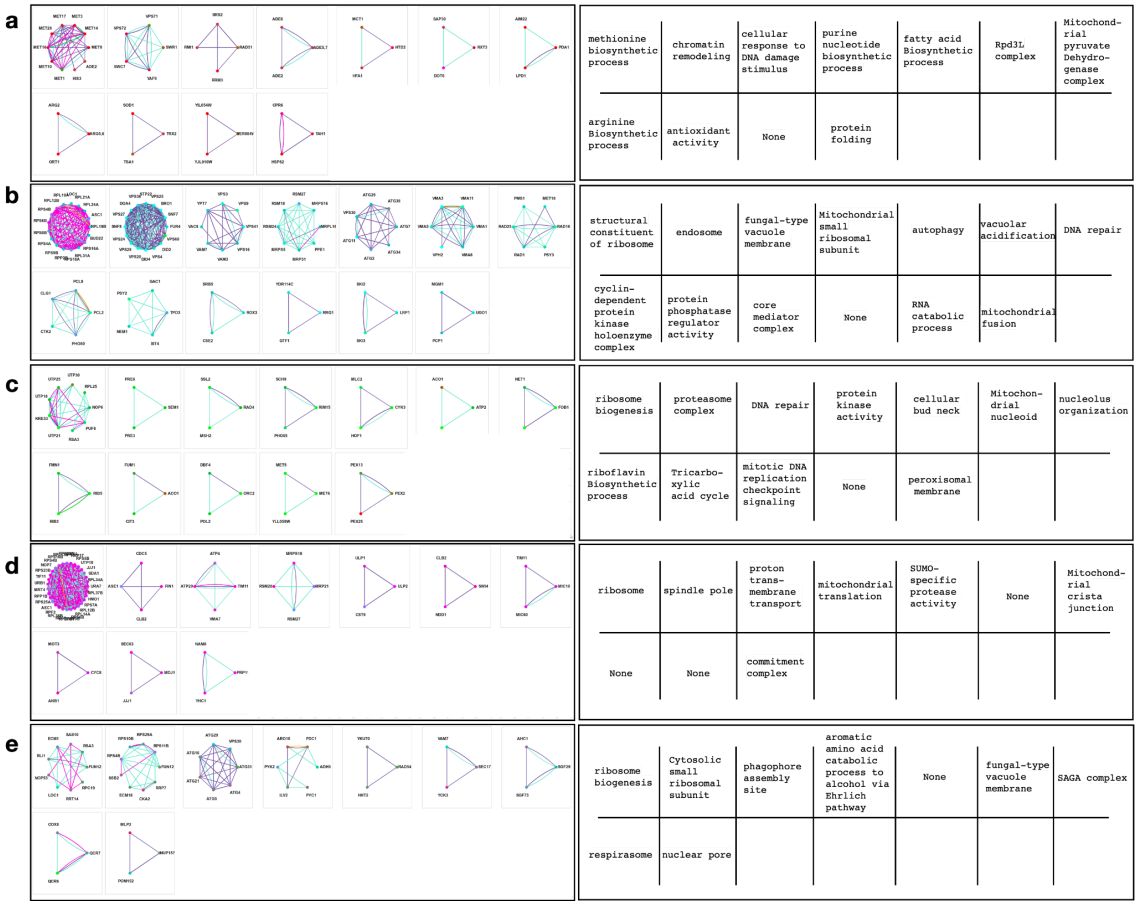


Supplementary Figure S3. Raw, uncorrected and corrected CRI-SPA data. a) Raw fluorescent images of screening plates produced by CRI-SPA for the YKO screen plate 2 (top) and the OEx screen (bottom). **b)** Uncorrected (top) and corrected (bottom) colony area and fluorescence data for YKO screen plate 2. Dots on the corrected plates mark empty spots and data points removed from our analysis. **c)** Distributions of the uncorrected (top) and corrected (bottom) data for each plate in the screen. Each colour is the distribution of data obtained for one plate.

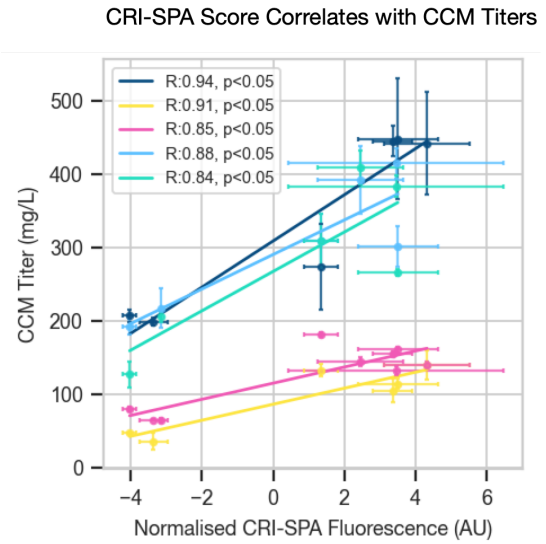


Supplementary Figure S4. STRING Clusters for 250 most extreme gene groups. Protein-protein interactions networks for the highest and lowest scoring genes in the YKO

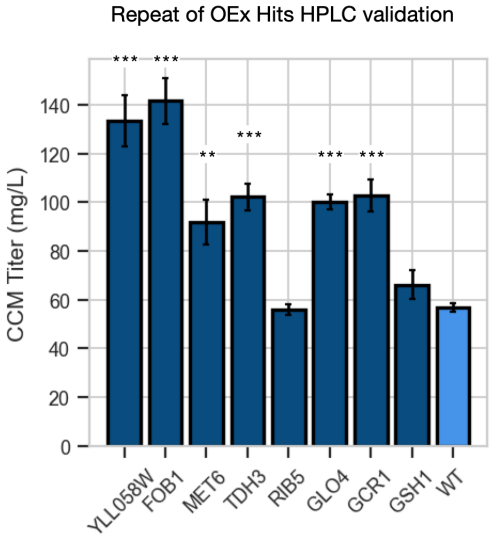
and the OEx screens. For the centre plot, each point is the mean fluorescence of a gene in the KO screen (x-axis) and the OEx screen (y-axis). Each gene was colour coded according to its x-y coordinate to map the effects of its KO and OEx on CCM synthesis in the rest of the figure. Graph nodes represent genes and are coloured according to their KO-OEx scores. For example, the top left graph represents the least fluorescent genes in the KO screen. Edges are STRING interactions of high confidence and coloured according to the bottom legend.



Supplementary Figure S5. K-clique clusters extracted from STRING networks and annotated with GOEA. a) All K-clique gene clusters extracted from the gene networks shown in **Suppl. Fig. S4**. **a)**Top fluorescent group in KO screen, **b)** Bottom fluorescent group in KO screen. **c)** Top fluorescent group in OEx screen. **d)** Bottom fluorescent group in OEx screen. **e**) Cliques extracted for a STRING network made for 250 randomly selected genes. Left panels show the cliques, right panels show representative GO terms enriched in the corresponding clique.



Supplementary Figure S6. CRI-SPA fluorescence score correlates with CCM titers in liquid culture. Scatter plot of fluorescences measured in the CRI-SPA screen and of CCM titers for top and bottom KO hits cultivated in SC medium for 72h. In the x-axis, each point is the mean fluorescence of three or more independent CRI-SPA colonies. In the y-axis, each point is the mean of three experimental replicates. Each colour represents a biological replicate performed on a different day. For both axes, the error bar is the standard deviation. R is the Pearson Correlation Coefficient between fluorescence intensities and CCM titers. AU = arbitrary units.



Supplementary Figure S7. Biological replicates of experiment presented in Figure 5. CCM titers for OEx top hits cultivated for 72h in liquid SC medium. Each bar represents the

mean of three experimental replicates. Significant difference from WT is assessed with a two-sided t-test and indicated by: *<0.05, **<0.01, ***<0.001.

Supplementary Table S1. Sourcing of parts used to build the CCM cassette.

Part name	Primer Fw	Primer Rv	Description	Source/template	Reference
pCYC_benO::yEGFP::tCyc1	CCM001	CCM002	GFP expression cassette under BenM activated promoter	pMeLS0025	Skjoedt et al
pTDH2	CMM005	CCM006	*S. cerevisiae* promoter	BY4741 genome	NA
ScTkl1	CMM007	CCM008	*S. cerevisiae TKL1*	pCFB1237	Skjoedt et al
tTDH3	CCM009	CCM010	*S. cerevisiae* terminator	BY4741 genome	NA
pFBA1	CCM011	CCM012	*S. cerevisiae* promoter	BY4741 genome	NA
PaAroZ	CCM013	CCM014	*Podospora anserina's* dehydroshikimate dehydratase	pCFB1239	Skjoedt et al
tHis5	CCM015	CCM016	*S. cerevisiae* terminator	BY4741 genome	NA
pTPI1	CCM017	CCM018	*S. cerevisiae* promoter	BY4741 genome	NA
NTC_R	CCM019	CCM020	Nourseothricin resistance	pCfB2193	Stovicek et al.
tLSC2	CCM021	CCM022	*S. cerevisiae* terminator	BY4741 genome	NA
pCYC1	CMM023	CMM024	*S. cerevisiae* promoter	BY4741 genome	NA
CaCatA	CMM025	CMM026	*Candida albicans*' catechol 1,2-dioxygenase	pCFB1239	Skjoedt et al
tCPS1	CMM027	CMM028	*S. cerevisiae* terminator	BY4741 genome	NA
pTDH3:KpAroY.B	CMM029	CMM030	*Klebsiella pneumonia* protocatechuic acid decarboxylase unit B under *S. cerevisiae's* pTdh3 promoter	pCfB1241	Skjoedt et al
pTDH3:KpAroY.B(P146T)	CMM029	CMM030	Optimised *Klebsiella pneumonia* protocatechuic acid decarboxylase unit B under *S. cerevisiae's* pTdh3 promoter	Gblock	ED Jensen 2021
tVPS13	CMM031	CMM032	*S. cerevisiae* terminator	BY4741 genome	NA
pTef1:KpAroyC.iso	CMM033	CMM034	*Klebsiella pneumonia* protocatechuic acid decarboxylase unit C under *S. Cerevisiae's* pTef1 promoter	pCFB1241	Skjoedt et al
tIDP1	CMM035	CMM036	*S. cerevisiae* terminator	BY4741 genome	NA
pCCW12	CMM037	CMM038	*S. cerevisiae* promoter	BY4741 genome	NA
KpAroY.D	CCM039	CCM040	*Klebsiella pneumonia* protocatechuic acid decarboxylase unit D	pCfB1237	Skjoedt et al
tPRM9	CCM041	CCM042	*S. cerevisiae* terminator	BY4741 genome	NA

BenM MP17_D08	CCM043	CCM044	Optimised version of BenM biosensor	pTS-117	Snoek 2020

Supplementary Table S2 Gene fluorescent ranking for the KO and OEx Screens

Find online

Supplementary Table S3. Gene hits selected for further testing.

Hit Name	Library	Normalised Fluorescence Score (Rank in Library)	Function
GCR1	OEx	3.39 (32th)	Transcriptional activator of genes involved in glycolysis
MET6	OEx	5.85 (8th)	cobalamin-independent methionine synthase
GLO4	OEx	1.62 (148th)	mitochondrial glyoxalase II, catalyzes the hydrolysis of s-d-lactoylglutathione into glutathione and d-lactate
GSH1	OEx	2.85 (43th)	catalyzes the first step in glutathione (gsh) biosynthesis
TDH3	OEx	2.00 (96th)	glyceraldehyde-3-phosphate dehydrogenase (GAPDH)
FOB1	OEx	5.60 (10th)	nucleolar protein that binds the rDNA replication fork barrier site
RIB5	OEx	4.53 (23th)	catalyzes the last step of the riboflavin biosynthesis
YLL058W	OEx	4.64 (21th)	Inefficient homocysteine synthase
LPD1	KO	3.45 (29th)	lipoamide dehydrogenase component (e3) of the pyruvate dehydrogenase
LYS14	KO	3.35 (32nd)	transcriptional activator regulating lysine biosynthesis
MET14	KO	3.53 (24rd)	adenylylsulfate kinase; required for sulfate assimilation and involved in methionine metabolism
PDA1	KO	4.32 (11th)	e1 alpha subunit of the pyruvate dehydrogenase (pdh) complex
AIM22	KO	2.44 (59th)	putative lipoate-protein ligase; required along with lip2 and lip5 for lipoylation of lat1p and kgd2p
ARG5,6	KO	3.50 (256h)	catalyzes the 2nd and 3rd step in arginine biosynthesis
VPS20	KO	-3.36 (4586th)	part of the endosomal sorting complex required for transport of transmembrane

			proteins into the multivesicular body pathway to the lysosomal/vacuolar lumen
ADE12	KO	-3.15 (4577th)	catalyzes the first step during purine nucleotide biosynthesis
SNF7	KO	-4.03 (4592nd)	involved in the sorting of transmembrane proteins into the multivesicular body (mvb) pathway

Supplementary Table S4 Strains used in this study

Publication name	**Parent**	**Genotype**	**Source**
CRISPA-donor strain mat-alpha	-	*MATα* CEN$_{1\text{-}16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)*	(Cachera et al., 2023)
CRISPA-donor strain mat-a	-	MATa CEN1-16::pGAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9-tCYC1-loxP-KlLEU2-loxP *KlURA3(dw XII-5 IS)*	(Cachera et al., 2023)
BY4741	-	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	(Baker Brachmann et al., 1998)
CD1	CRISPA-donor strain mat-alpha	*MATα* CEN$_{1\text{-}16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS) XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B**::AroY.D:: NrsR: CaCatA::Tkl1*	This study
CD2	CRISPA-donor strain mat-alpha	*MATα* CEN$_{1\text{-}16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS) XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.D**:: NrsR:: CaCatA::Tkl1*	This study
Sc_CCM052	CRISPA-donor strain mat-a	*MATa CEN1-16::pGAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9-tCYC1-loxP-KlLEU2-loxP KlURA3(dw XII-5 IS) XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B**::AroY.D:: NrsR: CaCatA::Tkl1*	This study
Sc_CCM053	CRISPA-donor strain mat-a	*MATa CEN1-16::pGAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9-tCYC1-loxP-KlLEU2-loxP KlURA3(dw XII-5 IS) XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.Ciso::::AroY.D**:: NrsR:: CaCatA::Tkl1*	This study
CD3	Sc_CCM052	*MATa CEN1-16::pGAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9-tCYC1-loxP-KlLEU2-loxP KlURA3(dw XII-5 IS) XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::AroY.B**::AroY.Ciso:::AroY.D**:: **KanR**:: CaCatA::Tkl1*	This study
CD4	Sc_CCM053	*MATa CEN1-16::pGAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5 X-3::pTEF1-SpCas9-tCYC1-loxP-KlLEU2-loxP KlURA3(dw XII-5 IS)*	This study

		*XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T:**::AroY.Ciso::: AroY.D:: **KanR**: CaCatA::Tkl1*	
BY0	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)*	This study
BY1	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B**::AroY.Ciso::AroY.D:: NrsR: CaCatA::Tkl1*	This study
BY2	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.Ciso:::AroY.D**:: NrsR:: CaCatA::Tkl1	This study
BY2-Aro4Δ	BY2	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 **Aro4Δ::KanMX** XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.Ciso::AroY.D**:: NrsR:: CaCatA::Tkl1	This study
BY2-Aro2Δ	BY2	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 **Aro2Δ::KanMX** XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.Ciso:AroY.D**:: NrsR:: CaCatA::Tkl1	This study
OExHit_GCR1	BY4741 (pHO38)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 **XII-5up**::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T**::AroY.Ciso:HomoB:AroY.D:KanR:: CaCatA::Tkl1::HomoC:Tef2p**GCR1**	This study
OExHit_MET6	BY4741 (pHO38)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XII-5up::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::AroY.B_P146T::AroY.Ciso:HomoB:AroY.D:KanR:: CaCatA::Tkl1::HomoC:TEF2p:**MET6**	This study
OExHit_:GLO4	BY4741 (pHO38)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XII-5up::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::AroY.B_P146T::AroY.Ciso:HomoB:AroY.D:KanR:: CaCatA::Tkl1::HomoC:TEF2p**:GLO4**	This study
OExHit_GSH1	BY4741 (pHO38)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XII-5up::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::AroY.B_P146T::AroY.Ciso:HomoB:AroY.D:KanR:: CaCatA::Tkl1::HomoC:TEF2p:**GSH1**	This study
OExHit_TDH3	BY4741 (pHO38)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XII-5up::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::AroY.B_P146T::AroY.Ciso:HomoB:AroY.D:KanR:: CaCatA::Tkl1::HomoC:TEF2p:**TDH3**	This study
OExHit_FOB1	BY4741 (pHO38)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XII-5up::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::AroY.B_P146T::AroY.Ciso:HomoB:AroY.D:KanR:: CaCatA::Tkl1::HomoC:TEF2p:**FOB1**	This study
OExHit_RIB5	BY4741 (pHO38)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XII-5up::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::AroY.B_P146T::AroY.Ciso:HomoB:AroY.D:KanR:: CaCatA::Tkl1::HomoC:TEF2p:**RIB5**	This study

OExHit_YLL058W	BY4741 (pHO38)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 XII-5up::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::AroY.B_P146T::AroY.Ciso:HomoB:AroY.D:KanR:: CaCatA::Tkl1::HomoC:TEF2p:**YLL058W**	This study
KOhit_Pda1	**Generated with CD2 x YKO CRI-SPA screen**	*MATα* CEN$_{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1* *RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)* *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.D**::NrsR:: CaCatA::Tkl1* **Pda1Δ**	This study
KOhit_ Arg5,6	**Generated with CD2 x YKO CRI-SPA screen**	*MATα* CEN$_{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1* *RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)* *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.D**::NrsR:: CaCatA::Tkl1* **Arg5,6Δ:KanMX**	This study
KOhit_Lys14	**Generated with CD2 x YKO CRI-SPA screen**	*MATα* CEN$_{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1* *RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)* *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.D**::NrsR:: CaCatA::Tkl1* **Lys14Δ:KanMX**	This study
KOhit_Met14	**Generated with CD2 x YKO CRI-SPA screen**	*MATα* CEN$_{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1* *RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)* *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.D**::NrsR:: CaCatA::Tkl1* **Met14Δ:KanMX**	This study
KOhit_Lpd1	**Generated with CD2 x YKO CRI-SPA screen**	*MATα* CEN$_{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1* *RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)* *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.D**::NrsR:: CaCatA::Tkl1* **Lpd1Δ:KanMX**	This study
KOhit_Snf7	**Generated with CD2 x YKO CRI-SPA screen**	*MATα* CEN$_{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1* *RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)* *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.D**::NrsR:: CaCatA::Tkl1* **Snf7Δ:KanMX**	This study
KOhit_ Ade12	**Generated with CD2 x YKO CRI-SPA screen**	*MATα* CEN$_{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1* *RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)* *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.D**::NrsR:: CaCatA::Tkl1* **Ade12Δ:KanMX**	This study
KOhit_Vps20	**Generated with CD2 x YKO CRI-SPA screen**	*MATα* CEN$_{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1* *RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)* *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.D**::NrsR:: CaCatA::Tkl1* **Vps20Δ:KanMX**	This study

KOhit_Aim22	**Generated with CD2 x YKO CRI-SPA screen**	*MATα* CEN$_{1\text{-}16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1* *RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)* *XII-5::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::**AroY.B_P146T::AroY.D**::NrsR:: CaCatA::Tkl1* **Aim22Δ:KanMX**	This study

Supplementary Table S5 Plasmids used in this study

Name	Backbone	Description	Type
pCCM001	pCFB2909	pCYC_benO::yEGFP::Cyc1t	CCM Cassette Assembly Round 1
pCCM003	pCFB2909	pTDH2::Tkl1:TDH3 t(bidir)	CCM Cassette Assembly Round 1
pCCM004	pCFB2909	pFBA1::PaAroZ::His5t	CCM Cassette Assembly Round 1
pCCM005	pCFB2909	pTPI1::NrsR::Lsc2t	CCM Cassette Assembly Round 1
pCCM006	pCFB2909	pCyc1:CaCatA:CPS1t	CCM Cassette Assembly Round 1
pCCM007	pCFB2909	pTDH3:KaroYB:VPS13t	CCM Cassette Assembly Round 1
pCCM008	pCFB2909	pTEF1:KaroY.Ciso:IDP1t	CCM Cassette Assembly Round 1
pCCM009	pCFB2909	pCCW12:AroYD:tPRM9	CCM Cassette Assembly Round 1
pCCM010	pCFB2909	tADH1:BenM(P17_D08):pRev1M	CCM Cassette Assembly Round 1
pCCM011	pCFB2909	pTDH3:KaroYB(P146T):VPS13t	CCM Cassette Assembly Round 1
pCCM012	pCCM021	XII-5up::PaAroZ::AroY.B:: AroY.Ciso:HomoB	CCM Cassette Assembly Round 2 for pathway alone

pCCM013	pCCM021	XII-5up::PaAroZ::AroY.B_P146T:: AroY.Ciso:HomoB	CCM Cassette Assembly Round 2 for pathway alone
pCCM014	pCCM023	HomoA::PaAroZ::AroY.B:: AroY.Ciso:HomoB	CCM Cassette Assembly Round 2 for pathway + sensor
pCCM015	pCCM023	HomoA::PaAroZ::AroY.B_P146T:: AroY.Ciso:HomoB	CCM Cassette Assembly Round 2 for pathway + sensor
pCCM016	pCCM022	HomoB: AroY.D:: NrsR:CaCatA::Tkl1::XII-5dwn	CCM Cassette Assembly Round 2 for pathway alone and for pathway + sensor
pCCM018	pCCM025	XII-5up::yEGFP::BenM(MP17_D08):XII-5dwn	CCM Cassette Assembly Round 2 Sensor alone
pCCM020	pCCM024	XII-5up::yEGFP::BenM(MP17_D08):HomoA	CCM Cassette Assembly Round 2 for pathway + sensor
pCCM021	pCFB2909	XII-5Up:clone:HomoB	Homology Backbone
pCCM022	pCFB2909	HomoB:clone:XII-5Down	Homology Backbone
pCCM023	pCFB2909	HomoA:clone:HomoB	Homology Backbone
pCCM024	pCFB2909	XII-5Up:clone:HomoA	Homology Backbone
pCCM026	pCCM023	Aro2:KanMX KO	Used to KO Aro2
pCCM027	pCCM023	Aro4:KanMX KO	Used to KO Aro4
pCCM039	pCCM022	HomoB:clone:HomoC	Used for OEx strains cloning
pCCM040	pCCM022	HomoC::clone::XII5	Used for OEx strains cloning
pCCM042	pCCM021	XII-5up::pCYC_BenO:yEGFP::BenM(MP17_D08)::PaAroZ::AroY.B_P146T::AroY.Ciso:HomoB	Used for OEx strains cloning
pCCM043	pCCM039	HomoB: AroY.D:KanR:: CaCatA::Tkl1::HomoC	Used for OEx strains cloning

pCCM044	pCCM021	Tp1_P:KanR:LSC2_T	Used to switch the NTC resistance on to Kan resistance in CD strains
pCCM055	pCCM040	HomoC:: pTEF2::GCR1 ::XII-5Dw	Used for OEx strains cloning
pCCM056	pCCM040	HomoC:: pTEF2::MET6 ::XII-5Dw	Used for OEx strains cloning
pCCM057	pCCM040	HomoC:: pTEF2::GLO4 ::XII-5Dw	Used for OEx strains cloning
pCCM058	pCCM040	HomoC:: pTEF2::GSH1 ::XII-5Dw	Used for OEx strains cloning
pCCM059	pCCM040	HomoC:: pTEF2::TDH3 ::XII-5Dw	Used for OEx strains cloning
pCCM061	pCCM040	HomoC:: pTEF2::FOB1 ::XII-5Dw	Used for OEx strains cloning
pCCM062	pCCM040	HomoC:: pTEF2::RIB5 ::XII-5Dw	Used for OEx strains cloning
pCCM063	pCCM040	HomoC:: pTEF2::YLL058W ::XII-5Dw	Used for OEx strains cloning
pHO029		pSNR52-gRNA(XII-5)-tSUP4	CRI-SPA vector with
pHO38		CRISPR-Cas9 (Hyg)	
pCFB3050		pSNR52-gRNA(XII-5)-tSUP4	

Supplementary Table S6 Primers used in this study

Primer Name	Bases	Tm (c)	User overhang	Template
CCM.001	CGTGCGAUGATCCAGGCAACTTTAGTGC	52.81	GV1	pMeLS0025(pCYC_benO::yEGFP::Cyc1t)
CCM.002	CACGCGAUTTCAGCTGACGCGATCT	50.11	GV2	pMeLS0025(pCYC_benO::yEGFP::Cyc1t)
CCM.003	CACGCGAUGAATGCGGCCGCTT	50.01	GV2	pMeLS0076(Rev1p::BenM (H110R, F211V, Y286N))
CCM.004	CGTGCGAUAGCGGATAACAATTTCACACA	51.41	GV1	pMeLS0076(Rev1p::BenM (H110R, F211V, Y286N))

CCM.005	AAGAGGGCUTTTGTTTTGTTTGTTTGTGTGATGAATTTAA	53.92	U1	pTDH2
CCM.006	CGTGCGAUATCTAGATCAGAGGGTGGT	50.32	GV1	pTDH2
CCM.007	AGCCTGTGUTCAGAAAGCTTTTTTCAAAGGA	49.14	U2	pCFB1237_ScTkl1
CCM.008	AGCCCTCTUACAATGACTCAATTCACTGACATTGAT	53.72	U1	pCFB1237_ScTkl1
CCM.009	CACGCGAUTCTGCAGGTAGGGAAAGA	51.07	GV2	ANTE113_Tdh3bidir
CCM.010	ACACAGGCUGTGAATTTACTTTAAATCTTGCATTTAA	50.04	U2	ANTE113_Tdh3bidir
CCM.011	AAGAGGGCuTTTGAATATGTATTACTTGGTTATG	48.18	U1	FBA1p
CCM.012	CGTGCGAuACTGGTAGAGAGCGACTTT	50.78	GV1	FBA1p
CCM.013	AGCCTGTGuTCACAAAGCAGCTGACAAAG	51.02	U2	pCFB1239_PaAroZ
CCM.014	AGCCCTCTuACAATGCCATCCAAGTTGG	52.12	U1	pCFB1239_PaAroZ
CCM.015	ACACAGGCuATAGATTAATTTAAACAGTATATGTACAGTTT	51.49	U2	His5t
CCM.016	CACGCGAuGTAACAATATCATGAGACCTTTTATA	49.64	GV2	His5t
CCM.017	AAGAGGGCuTTTTAGTTTATGTATGTGTTTTTTGTAGT	51.25	U1	TPI1p
CCM.018	CGTGCGAuAAGGATGAGCCAAGAATAAGG	51.72	GV1	TPI1p
CCM.019	AGCCCTCTuACCATGGGTACCACTCT	51.25	U1	pCfB2193_NatMX
CCM.020	AGCCTGTGuTTAGGGGCAGGGCA	51.63	U2	pCfB2193_NatMX
CCM.021	CACGCGAuAAAATTAAAAAAAAAAAAAAGAAATTTTTTCC	48.92	GV2	LSC2t
CCM.022	ACACAGGCuGCTTCTCGAGAAAAACAAAAGAGTTA	53.07	U2	LSC2t
CCM.023	CGTGCGAuCAGCATTTTCAAAGGTGT	45.87	GV1	pCYC1
CCM.024	AAGAGGGCuTATTAATTTAGTGTGTGTATTTGTGTT	50.24	U1	pCYC1
CCM.025	AGCCCTCTuACAATGTCCCAAGCTTTCA	50.21	U1	pCFB1239CaCatA

CCM.026	AGCCTGTGuTTACAACTTGATTTCAGCATCTTG	50.74	U2	pCFB1239CaCatA
CCM.027	ACACAGGCuGCGCAATGATTGAATAGTCAAA	50.74	U2	CPS1t
CCM.028	CACGCGAuGATTTGACACTTGATTTGACACTTCTTT	53.48	GV2	CPS1t
CCM.029	AGCCTGTGuTCATTCAATTTCTTGAGCGAATTG	50.87	U2	pCfB1241_TDH3p:KpAroY.B
CCM.030	CGTGCGAuGCTATAAAAAACACGCTTTTTCA	50.16	GV1	pCfB1241_TDH3p:KpAroY.B
CCM.031	CACGCGAuGCGCGCTGCGGA	51.74	GV2	VPS13t
CCM.032	ACACAGGCuTCACATATGAAAGTATATACCCG	50.63	U2	VPS13t
CCM.033	CGTGCGAuGCACACACCATAGCTTCAA	51.31	GV1	pCFB1241_Tef1p_KpAroyC.iso
CCM.034	AGCCTGTGuTTACTTAGCGGAACCTTGATT	51.33	U2	pCFB1241_Tef1p_KpAroyC.iso
CCM.035	ACACAGGCuTCGAATTTACGTAGCCCAATCTA	54.23	U2	IDP1t
CCM.036	CACGCGAuGATGGTAATGATCCGAACTTGG	53.28	GV2	IDP1t
CCM.037	AAGAGGGCuTATTGATATAGTGTTTAAGCGAATGA	50.87	U1	CCW12p
CCM.038	CGTGCGAuAAAGAAACTTAATACGTTATGCC	49.65	GV1	CCW12p
CCM.039	AGCCCTCTuACAATGATCTGTCCAAGATGC	50.99	U1	pCfB1237_KpAroY.D
CCM.040	AGCCTGTGuTTATCTCTTATCTTCTGGCAATAATG	51.11	U2	pCfB1237_KpAroY.D
CCM.041	ACACAGGCuACAGAAGACGGGAGACAC	51.44	U2	PRM9t
CCM.042	CACGCGAuATTTTCAACATCGTATTTTCCGA	50.34	GV2	PRM9t
CCM.043	CGTGCGAuGAATGCGGCCGCTT	50.01	GV1	pMeLS0076(Rev1p::BenM (H110R, F211V, Y286N))
CCM.044	CACGCGAuAGCGGATAACAATTTCACACA	51.41	GV2	pMeLS0076(Rev1p::BenM (H110R, F211V, Y286N))
CCM.045	AAGAGGGCuctcgctgaggacttaatgc		U1	pCFB2909

CCM.046	AGCCCTCTugaatgcgtgcgatcgcgtgcattcACTGAATACAGGCTAGTGAATGCTCTCGGTCTTGTACGTGTGCGATTGACagattaagacctcagcgc		U1	pCFB2909
CCM.047	AGCCCTCTugaatgcgtgcgatcgcgtgcattcagtgctgaggcattaatgc		U1	pCFB2909
CCM.048	AAGAGGGCuGTCAATCGCACACGTACAAGACCGAGAGCATTCACTAGCCTGTATTCAGTtgattaaaccctcagcgcg		U1	pCFB2909
CCM.049	AAGAGGGCuGTTGTCGTGCCACCGTTACCTTCGTGAGTATTCAATCTCATAGCCGAGTAtgattaaaccctcagcgcg		U1	pCFB2909
CCM.050	AGCCCTCTugaatgcgtgcgatcgcgtgcattcTACTCGGCTATGAGATTGAATACTCACGAAGGTAACGGTGGCACGACAACcagattaagacctcagcgc	102	U1	pCFB2909
CCM.012	CGTGCGAuACTGGTAGAGAGCGACTTT	50.78	GV1	pCCM004(pFBA1:PaAroZ:tHis5)
CCM.051	AAGAGGGCuGTAACAATATCATGAGACCTTTTATAG	50.67	U1	pCCM004(pFBA1:PaAroZ:tHis5)
CCM.052	AGCCCTCTuATAAAAAACACGCTTTTTCAGTTC	50.2	U1	pCCM007(pTDH3:KpAroyB:tVPS13)
CCM.053	AGCCTGTGuGCGCGCTGCGGA	51.74	U2	pCCM007(pTDH3:KpAroyB:tVPS13)
CCM.054	ACACAGGCuGCACACACCATAGCTTCA	50.54	U2	pCCM008(pTEF1:Aroy.Ciso:tIDP1)
CCM.036	CACGCGAuGATGGTAATGATCCGAACTTGG		GV2	pCCM008(pTEF1:Aroy.Ciso:tIDP1)
CCM.038	CGTGCGAuAAAGAAACTTAATACGTTATGCC		GV1	pCCM009(pCCW12:AroYD:tPRM9)
CCM.055	AAGAGGGCuATTTTCAACATCGTATTTTCCGA	50.34	U1	pCCM009(pCCW12:AroYD:tPRM9)
CCM.056	AGCCCTCTuTAAGGATGAGCCAAGAATAAGG	52.15	U1	pCCM005(pTPI1:NrsR:LSC2t)
CCM.057	AGCCTGTGuAAAATTAAAAAAAAAAAAAAGAAATTTTTTCCA	50.07	U2	pCCM005(pTPI1:NrsR:LSC2t)
CCM.058	ACACAGGCuTCAGCATTTTCAAAGGTGTGT	51.15	U2	pCCM006(pCyc1:CaCatA:CPS1t)
CCM.059	AGGTGGCAuGATTTGACACTTGATTTGACACTT	50.4	U3	pCCM006(pCyc1:CaCatA:CPS1t)

CCM.060	ATGCCACCuATCTAGATCAGAGGGTGGT	50.32	U3	pCCM003(pTDH2:Tkl1:TDH3t)
CCM.009	CACGCGAuTCTGCAGGTAGGGAAAGA	51.07	GV2	pCCM003(pTDH2:Tkl1:TDH3t)
CCM.001	CGTGCGAuGATCCAGGCAACTTTAGTGC	52.81	GV1	pCCM001(pCYC_benO:yEGFP:Cyc1t)
CCM.061	AAGAGGGCuTTCAGCTGACGCGATCT	50.11	U1	pCCM001(pCYC_benO:yEGFP:Cyc1t)
CCM.062	CACGCGAuGCGACCTCATGCTATACC	51.71	GV2	pCCM002(tADH1::BenM::pRev1) and pCCM010(tADH1::BenM(MP17_D08)::pRev1
CCM.063	AGCCCTCTuAGCGGATAACAATTTCACACA	51.41	U1	pCCM002(tADH1::BenM::pRev1)
CCM.063B	AGCCCTCTuTTCTTAGGCACAACAATATTTATAAA		U1	pCCM010(tADH1::BenM(MP17_D08)::pRev1
CCM.064	CGTGCGAuTTACCAATTTGGTGGTTCAG	50.21	GV1	pMeLS0076(Rev1p::BenM_MP17_D08)
CCM.065	CACGCGAuTTCTTAGGCACAACAATATTTATAAA	50.1	GV2	pMeLS0076(Rev1p::BenM_MP17_D08)

References

Baker Brachmann, C., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., Boeke, J.D., 1998. Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast 14, 115–132.

Cachera, P., Olsson, H., Coumou, H., Jensen, M.L., Sánchez, B.J., Strucko, T., van den Broek, M., Daran, J.-M., Jensen, M.K., Sonnenschein, N., others, 2023. CRI-SPA: a high-throughput method for systematic genetic editing of yeast libraries. Nucleic Acids Res. gkad656.