

**Institut für Biochemie und Biologie**

Arbeitsgruppe Cell2Fab, Synthetische Biologie (Prof. Dr. Bernd Müller-Röber)

---

**Plant-derived transcription factors and their application for synthetic biology approaches  
in *Saccharomyces cerevisiae***

Dissertation  
zur Erlangung des akademischen Grades  
"doctor rerum naturalium"  
(Dr. rer. nat.)  
in der Wissenschaftsdisziplin "Synthetische Biologie"

eingereicht an der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der Universität Potsdam

von  
Gita Naseri

Potsdam, den 15.05.2018

Published online at the

Institutional Repository of the University of Potsdam:

URN urn:nbn:de:kobv:517-opus4-421514

<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-421514>

# TABLE OF CONTENTS

---

|  |    |
|--|----|
| ACKNOWLEDGEMENT.....   | i  |
| 1 INDEX OF ABBREVIATIONS .....   | 1  |
| 2 SUMMARY.....   | 3  |
| 3 INTRODUCTION.....  | 8  |
| 3.1 Synbio: Interrogating Organizational Principles of Living Systems .....  | 8  |
| 3.2 Yeast: Microbial Chassis Platform.....   | 9  |
| 3.3 Constructing the Toolbox: Genetic Parts, Tools and Circuits .....  | 9  |
| 3.4 Challenges for Synbio: Hard Truths of Synbio Labs.....   | 10 |
| 3.5 Plant Transcription Factors and Promoter Pairs.....  | 11 |
| 3.6 Artificial Transcription Factors and Promoter Pairs .....  | 11 |
| 3.7 Episomal <i>versus</i> Integration .....   | 12 |
| 3.8 CRISPR/Cas-based Technology: Synbio Gemstone .....   | 13 |
| 3.9 Metabolic Engineering: White Biotechnology .....   | 15 |
| 3.10 Carotenoids .....   | 16 |
| 3.11 Naringenin .....  | 16 |
| 3.12 Rational Techniques for Metabolic Engineering: Classical Methods.....   | 17 |
| 3.13 Non-rational Techniques for Metabolic Engineering.....  | 17 |
| 3.14 Biosensor: Cell Factory Lighter .....   | 18 |
| 3.15 A Vision of the Future of Synbio .....  | 19 |
| 3.16 Goals of this Study .....   | 20 |
| 4 Plant-derived transcription factors for orthologous regulation of gene expression in the yeast <i>Saccharomyces cerevisiae</i> ..... | 22 |

## TABLE OF CONTENTS

---

|   |   |     |
|---|---|-----|
| 5 | COMPASS: Rapid combinatorial optimization of biochemical pathways based on artificial transcription factors .....   | 56  |
| 6 | CaPRedit: Genome editing using CRISPR-Cas9 and plant-derived transcriptional regulators for the redirection of flux through the FPP branch-point in yeast ..... | 121 |
| 7 | DISCUSSION .....  | 160 |
| 8 | REFERENCES.....   | 173 |
|   | SCIENTIFIC ACHIEVEMENTS .....   | 184 |
|   | DECLARATIONS .....  | 185 |
|   | DECLARATION OF INDEPENDENT WORK.....  | 185 |
|   | DECLARATION OF OWN CONTRIBUTIONS TO SINGLE PUBLICATIONS.....  | 186 |



## ACKNOWLEDGEMENT

---

### **ACKNOWLEDGEMENTS**

I would like to thank my supervisor, Prof. Dr. Bernd Mueller-Roeber, for giving me the opportunity to pursue my interests in synthetic biology, and also for all his kind help, discussions, advice and support. I would like to acknowledge my second supervisor, Prof. Dr. Katja Arndt and Dr. Katrin Messerschmidt for her support in the Cell2Fab group.

I would also like to appreciate Prof. Dr. Matias Zurbriggen and Prof. Dr. Alain Tessier for agreeing to be referees of this work.

I would also like to acknowledge Jessica Behrend, Lisa Rieper, and Karina Schulz for their help with experiments. Furthermore, I want to thank my colleagues and the whole Mueller-Roeber group.

I am particularly grateful to my husband, Parham Bahadornjad, for his love, encouragement, and patience throughout my study, which gave me the strength to achieve my goals.

Last, but not least, I want to express my deep appreciation to my parents, my sisters, and my brother for all their love.

Gita Naseri

15.05. 2018



## 1 INDEX OF ABBREVIATIONS

|                    |  |
|--------------------|--|
| aa                 | Amino acids  |
| AD                 | Activation domain  |
| ATc                | Anhydrotetracycline  |
| ATF                | Artificial transcription factor  |
| ATFP               | Artificial transcription factor and synthetic promoter containing the corresponding binding site |
| <i>A. thaliana</i> | <i>Arabidopsis thaliana</i>  |
| AU                 | Arbitrary Unit   |
| BLAST              | Basic local alignment search tool  |
| BS                 | Binding site   |
| bp                 | Base pairs   |
| Cas                | CRISPR-associated  |
| CCM                | Cis, <i>cis</i> -muconic acid  |
| CDS                | Coding DNA sequence  |
| CDW                | Cell dry weight  |
| CoA                | Coenzyme A   |
| CRISPR             | Clustered regularly interspaced short palindromic repeats  |
| DBD                | DNA-binding domain   |
| dCas9              | Dead Cas9  |
| ddH <sub>2</sub> O | double distilled water   |
| dsOligo            | Double-strand oligonucleotides   |
| DNA                | Deoxyribonucleic acid  |
| dNTP               | Deoxynucleotide triphosphate   |
| <i>E. coli</i>     | <i>Escherichia coli</i>  |
| GC-MS              | Gas chromatography-mass spectrometry   |
| GOI                | Gene of interest   |
| GMO                | Genetically modified organism  |
| HPLC               | High-performance chromatography  |
| HR                 | Homology region  |

## INDEX OF ABBREVIATIONS

---

|                       |   |
|-----------------------|---|
| IPTG                  | Isopropyl- $\beta$ -D-thiogalactopyranoside                                 |
| kb                    | Kilo base   |
| LB                    | Lysogeny broth  |
| LTR                   | LysR-type transcriptional regulator   |
| mRNA                  | Messenger RNA   |
| NG                    | Naringenin  |
| ORF                   | Open reading frame  |
| PAM                   | Protospacer adjacent motif  |
| P <sub>ATF</sub>      | IPTG-inducible promoter deriving artificial transcription factor expression |
| PBS                   | Phosphate buffered saline   |
| PCR                   | Polymerase chain reaction   |
| RNA                   | Ribonucleic acid  |
| qRT-PCR               | Quantitative real time PCR  |
| SC                    | Synthetic complete media  |
| <i>S. cerevisiae</i>  | <i>Saccharomyces cerevisiae</i>   |
| gRNA                  | Guide RNA   |
| SLiCE                 | Seamless Ligation Cloning Extract   |
| ssDNA                 | Single-strand DNA   |
| <i>T. baccata</i>     | <i>Taxus baccata</i>  |
| <i>T. cuspidate</i>   | <i>Taxus cuspidate</i>  |
| TALE                  | Transcription activator-like effector                                       |
| TAR                   | Transformation-associated recombination                                     |
| tetO <sub>2</sub>     | Tetracycline operator   |
| tetR                  | Tetracycline repressor  |
| TF                    | Transcription factor  |
| TU                    | Transcriptional unit  |
| UTR                   | Untranslated region   |
| UV                    | Ultra-violet  |
| <i>X. dendrorhous</i> | <i>Xanthophyllomyces dendrorhous</i>  |
| YAC                   | Yeast artificial chromosome   |
| WT                    | Wild type   |
| ZFP                   | Zinc finger protein   |

## 2 SUMMARY

*Saccharomyces cerevisiae* is a host organism often used for synthetic biology (synbio) projects. Engineering such microorganisms for the production of compounds involves iterations of design, build, and test phases, in which a metabolic design is implemented by genetic engineering and later tested. Presently, the build phase, in which DNA molecules (*i.e.* promoters and genes) are assembled into plasmids and/or the host organism's genome, is the most time-consuming step in such approaches. Here, novel tools and techniques were developed to accelerate the build phase.

Yeast constitutive promoters are commonly used to express heterologous enzymes in this species, but this is often metabolically burdensome for the cell. Therefore, orthogonal regulatory tools for the tight and specific control of heterologous enzyme expression are required. Recently, pairs of artificial transcription factors and synthetic target promoters (minimal promoter fused to binding site(s) of the ATF) (ATFPs) were developed as orthogonal regulatory modules. The expression of ATFs is usually regulated by an inducible promoter ( $P_{ATF}$ ), allowing control of the timing of ATF expression and subsequent binding to the synthetic promoter driving expression of a target gene.

Although ATFP systems provide an alternative to constitutive promoters, current methods for building ATFPs are labor-intensive, and the strength of ATF-mediated transcriptional activation is often low. Furthermore, increasingly complex synthesis tasks for engineered organisms require a wider range of transcriptional activation capacities. An additional challenge in eukaryotic hosts (*e.g.*, yeast) is that a distinct module (*e.g.* an ATFP) is required to control the expression of each gene in the engineered pathway. This is in contrast to the situation in prokaryotic systems where sets of genes (coding sequences) can be transcribed from a single regulatory unit, the operon. As yeast's transcriptional machinery is unable to support polycistronic gene expression, synthetic biologists focused on developing orthogonal ATFPs that can provide valuable tools for establishing gene control networks in the build phase.

Another major challenge in the build phase stems from a lack of prior knowledge about how to optimize metabolic flux toward the desired product; more specifically, what is the optimal level of each enzyme in a biosynthetic pathway? In some cases, lower flux can result in higher product yield because production of foreign pathway products at a high level can have a toxic effect on cells, leading to growth inhibition. In addition, laboratory and industrial strains have different genetic backgrounds, so metabolic flux optimization achieved in laboratory strains may not be easily transferable to industrial strains.

In the last few years, several novel approaches have been established in yeast to avoid these potential obstacles by optimizing the metabolic flux at the level of DNA, RNA, and protein through creating more combinations of metabolic engineering targets. Development of so-called combinatorial metabolic engineering approaches allowing modification of the host genome in a modular, parallel, and high-throughput manner are essential to produce a library of strains where its members vary with respect to the level of the final compound produced. Recently, combinatorial approaches implementing yeast constitutive promoters have been used to optimize pathway production at the transcriptional level in yeast. However, as discussed above, transcription systems based on constitutive yeast promoters have important limitations. Combinatorial approaches implementing inducible ATFPs for the regulation of gene expression in yeast have to our knowledge not yet been reported. Plant-derived ATFPs have the potential to form the basis of new transcription-based combinatorial expression tools and may therefore greatly speed up the build phase.

Another approach to improve heterologous biosynthetic pathway production is increasing the precursor and/or cofactor supplies in *S. cerevisiae* by redirecting endogenous metabolism toward the increased production of precursors that are central for the biosynthesis of wanted products. Combining two recently developed tools, *i.e.*, CRISPR/Cas9-mediated one-step multigene modification and inducible plant-derived ATFPs, provide new tools for rapidly engineering metabolic pathways for an improved production of wanted chemicals. Taken together, these methodological developments would likely greatly improve the efficiency of *S. cerevisiae* as a cell factory.

In this study, a new class of inducible ATFPs was developed for use in yeast on the basis of transcription factors (TFs) derived from plants. TFs typically contain two functional domains, namely the DNA-binding domain (DBD) that binds to the promoters of target genes, and the activation domain (AD) that activates transcription by interacting with the basal transcription machinery of the cell. Plant-derived ATFPs were generated by integrating ATFs and synthetic promoters containing binding sites for the ATF into the genome of *S. cerevisiae*. A wide range of transcriptional outputs was achieved using different combinations of plant TF families and artificial ADs. Different ADs, such as the herpes simplex virus protein VP16 AD, the yeast GAL4 AD, and the plant EDLL AD were used. For first time, it was reported that the plant EDLL motif can be successfully employed as a strong AD in yeast, and thereby it is a suitable AD to generate strong orthogonal ATFPs. The transcriptional output of several plant TF-EDLL fusions together with the synthetic promoter containing TF binding sites exceeded the transcriptional output of the strong yeast *TDH3* promoter by 6- to 10-fold.

## SUMMARY

---

However, for full exploitation of genetic circuit construction tools, transcription regulatory modules that can function independently of each other are needed. Therefore, it was examined whether two different plant-derived ATFPs, encoded on the same plasmid, can be expressed independently and can independently drive gene expression. More specifically, the expression of two ATFs encoded by either centromeric or episomal plasmids was characterized. It was shown that isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG)- and anhydrotetracycline (ATc)-inducible GAL1 promoters can be used for the independent transcriptional activation of two plant-derived ATFPs and that these two ATFPs can independently control the expression of two different fluorescent proteins.

As mentioned above, high-level heterologous gene expression is not always the best solution for metabolic engineering projects. Here, a COMbinatorial Pathway ASSEMBly (COMPASS) cloning approach was reported, in which the expression of plant-derived ATFPs is controlled by an IPTG-inducible *GAL1* promoter and plant-derived ATFPs control the expression of coding DNA sequences (CDSs). The approach relies on combinatorial cloning to generate all possible combinations between ATFP control modules and CDSs. The combinatorial libraries of ATFPs of varying strengths (weak, medium and strong) lead to the expression of each CDS at different levels. This diversity allows the optimization of pathway gene expression for maximal product output without *a priori* knowledge of the best combination of expression levels of the individual genes.

In order to be as tractable as possible, COMPASS was designed to allow combinatorial optimization of heterologous pathways on yeast artificial chromosomes (YACs). Additionally, to avoid segregational and/or structural instability of plasmid-based pathway expression, and because the integration of pathways into the chromosomal DNA often leads to higher yields compared to pathways on plasmids, COMPASS provides a setup for YAC integration into the genome. The CRISPR/Cas9 system employed by COMPASS allows for one-step integration of multiple cassettes into several distinct loci.

Here, COMPASS was applied to assembly pathways for  $\beta$ -carotene and  $\beta$ -ionone production, and then extended this by creating yeast cells jointly producing  $\beta$ -ionone and naringenin (NG), whereby the accumulation of NG was sensed at the single-cell level using a recently reported NG biosensor.  $\beta$ -Carotene and  $\beta$ -ionone are isoprenoids whose pigmented products are observable in yeast colonies, making them tractable model products for testing combinatorial approaches.  $\beta$ -Carotene is also the precursor of vitamin A and can be converted into many other secondary metabolites, including  $\beta$ -ionone, an aroma apo-carotenoid used in flavors and a key

intermediate in the synthesis of vitamin A. Moreover, carotenoids are used as food-coloring agents, antioxidants, in aquacultures, in cosmetics, and in the pharmaceutical industry. NG was also chosen as a COMPASS proof-of-concept product because it is a key intermediate in flavonoid production from tyrosine and phenylalanine. Flavonoids include a huge family of plant secondary metabolites that show a wide diversity of antioxidant and human health-related properties. Because of the value of these products, developing high-throughput methods to speed up yeast biofactory strain development via combinatorial approaches is an important goal.

In next effort, a CRISPR/Cas9- and plant-derived regulator-mediated genome editing approach (CaPRedit) were developed. Speeding up the strain modification and the large-scale changing of enzyme expression in desirable stage of growth, through implementing CRISPR/Cas9-mediated one-step multigene modification system and inducible superior plant-derived ATFPs (10-fold stronger than the yeast constitutive strong *TDH3* promoter), are two key features that CaPRedit is developed for. CaPRedit could be implemented for increased production of endogenous supplies through redirecting the yeast endogenous metabolic flux toward a key precursor of a desired heterologous product. Moreover, CaPRedit can also be implemented to increase the production of heterologous enzymes involved in a biosynthetic pathway engineered in *S. cerevisiae*. As a proof of principle, endogenous metabolites were redirected towards farnesyl diphosphate (FPP) production, a central precursor to nearly all isoprenoid products, including  $\beta$ -carotene and  $\beta$ -ionone. Using this approach, CaPRedit\_FPP 1.0 strain was generated, in which three genes, *tHMG1*, *ERG20*, and *GDH2*, were inducibly overexpressed under the control of strong plant-derived ATFPs. Notably, the production of  $\beta$ -carotene (one product upstream of  $\beta$ -ionone) in the CaPRedit\_FPP 1.0 strain was markedly improved compared to the production in a wild-type strain (by 4.3-fold) and the strain that had previously been optimized for FPP production using a combination of constitutive overexpression, constitutive down regulation, and gene deletion (by 1.3-fold).

In summary, in this thesis a panel of plant-derived ATFPs were developed and were implemented them in COMPASS, a combinatorial library of inducible ATFPs that provides a fast way to optimize biosynthetic pathway production in yeast synbio applications. Additionally, a strategy for speeding up strain modification was developed that allows low to high expression of enzymes in yeast via implementing a large collection of inducible plant-derived ATFPs and the CRISPR/Cas9 editing system; dubbed the new tool CaPRedit. The synbio toolbox presented here speeds up the build phase with the goal of moving the yeast *S. cerevisiae* from being a laboratory host to a general industrial biotechnology host for production of desired functionalities.





## 3 INTRODUCTION

### 3.1 Synbio: Interrogating Organizational Principles of Living Systems

Synbio denotes the design and fabrication of biological modules and systems that do not exist in the natural world, or the redesign of existing biological systems to create life from scratch. The term 'synthetic biology' was first used by Stéphane Leduc. In 'The Mechanism of Life', published in 1911, Leduc attempted to show that the origin of life is merely a chemical process.<sup>1</sup> Later, the term 'synbio' was used to describe the creation of synthetic molecules *in vivo* that mimic natural molecules.<sup>2</sup> Synthetic biologists aim to develop tools and modules that can contribute to the whole system in an independent manner and, therefore, the performance of an assembly can be predicted. Engineering is defined by the Oxford Dictionary as the branch of science and technology that is concerned with the design, building, and use of engines, machines, and structures. Therefore, synbio is a kind of genetic engineering. It is the combination of two principles, predictable tools and circuits and rapid prototyping,<sup>3</sup> which provides synbio with the ability to powerfully address the challenges in energy,<sup>4</sup> agriculture,<sup>5</sup> and human health.<sup>6</sup>

Synbio is an emerging field, in which the appearance of new developments, devices, and applications is accelerating. Synthetic biologists have already created tools to diagnose diseases like AIDS<sup>7</sup> and produce important molecules at low-cost such as a version of the drug artemisinin for poor people living in mosquito-ridden countries.<sup>8</sup> Moreover, recently scientists developed a method to use DNA to store video data, paving the way for the development of a molecular recorder that will sit inside living cells and collect data over time.<sup>9</sup> Synbio may also restore biodiversity and boost ecosystem productivity.<sup>10</sup>

Alongside the profound potential utility of synbio, there is fear that synthetic biologists are interfering with the theological concepts of creation, as created in the image of God, which makes synbio as a threat to human well-being.<sup>10</sup> Moreover, there are numerous open questions about the safety of synthetic organisms and the possibility of unforeseen consequences.<sup>11</sup> As with the fields of subatomic physics and nuclear energy, synbio development surely brings with it a number of threats. For example, in the wrong hands this technology is able to produce devastating biological weapons, and without proper safeguards it could lead to ecosystem destruction. One notable issue is the potential for accidental release of redesigned organisms. Will scientists be obligated to take legal responsibility for the unintended consequences of their discoveries and inventions? Which governmental body will be responsible for oversight and

enforcement and to what extent and with which tools will synthetic biologists be allowed to manipulate organisms? Questions such as these are profound and will require decades of dialog among many parts of society. Even before the 'rules of the game' are formalized, synbio practitioners must strive to maximize benefits for society and the environment while always minimizing the potential for harm. Overall, the public needs to get better at communicating what an amazing revolution synbio is, while at the same time scientists need to be open about doubts as they arise and the complexity with regards to possible impacts on social sustainability, maintenance of ecological diversity, and reversibility of decisions.

Could this emerging field help us deal with some of the toughest issues facing us, such as climate change, pollution and world hunger? Likely future will tell us.

### **3.2 Yeast: Microbial Chassis Platform**

Humans have been using yeasts in the production of food and alcoholic beverages for several millennia.<sup>12</sup> In the last few decades, yeast biotechnology has become a hot field, as advances in genomics, metabolic engineering, systems biology, and synbio have produced synergies enabling yeast to produce many high-value primary and secondary metabolites, enzymes, and pharmaceutical proteins.<sup>13-14</sup> Recently, Keasling *et al.* implemented synbio approaches to generate *Saccharomyces cerevisiae* producing artemisinic acid with a production capacity of 25 g/l, a much higher yield than has been reached in *E. coli*.<sup>15</sup> Yeast has several advantages for cost-effective molecular fabrication of valuable products such as (i) great tolerance to changes of environmental conditions such as temperature, pH and osmotic stress; (ii) formation of haploid cells offering single-gene, single-phenotype relationship, and diploid cells allowing robust growth and increased adaptation; (iii) post-translational modification machinery similar to that of other eukaryotes.<sup>16</sup>

### **3.3 Constructing the Toolbox: Genetic Parts, Tools and Circuits**

To undertake engineering projects two important aspects need to be considered: (i) the availability of biological tools and (ii) the possibility to easily assemble multi-component DNA constructs. Genetic sequences that perform needed functions are called the 'parts', and 'tools' combine parts to achieve more complex functions; for example a synthetic promoter is derived by fusing a minimal promoter and the binding site of a transcription factor (TF).<sup>17</sup> In the past decade, synthetic biologists have built a huge collection of tools for synbio applications. Recently, Tsai *et al.*<sup>18</sup> reviewed the important genetic tools, such as promoters, artificial

transcription factors (ATFs) and their target promoters, terminators,<sup>19</sup> sensors, and reporters,<sup>20</sup> for synbio projects in yeast. Circuits are assembled biological tools with specific biological functions such as promoter derived expression of a protein. Logical gates are a particularly powerful circuit type that translates perception of extracellular or intracellular signal into activation of cellular decision-making procedures and reprogramming the behavior of cell similar to logic gates in electrical circuits. Genetic devices such as RNA riboswitches, oscillators, amplifiers, and recorders are some examples.<sup>21</sup> Logical gates can be used to control the timely expression of genes so as to minimize energy and nutrient consumption and maximize production capacity.<sup>22</sup> The further development of genetic circuits with multiple controllable tools for more diverse logical gates would likely accelerate synbio advances.<sup>22</sup> Additionally, many synbio projects require cloning methods suitable for the assembly of complex DNA constructs.<sup>23-24</sup> The development of diverse assembly methods provides researchers with the possibility to choose and combine different approaches depending on the specific synbio challenge.<sup>25-26</sup> However, most of the established methods still need a well-established synbio laboratory with practical knowledge and experience in multi-part assemblies.

### **3.4 Challenges for Synbio: Hard Truths of Synbio Labs**

As all life is based on the same genetic code, synbio can provide a toolbox of reusable genetic parts and tools to form circuits with desired functions.<sup>25-26</sup> However, currently the synbio field is facing the reality of our very limited knowledge about the details of how life works.<sup>27-28</sup> Many of the parts are undefined and are incompatible when circuits get large and the process of constructing and testing them becomes more complex.<sup>3, 22, 29</sup> Variation in growth conditions can also affect the final function of circuits and force the organism for random mutation that might not be desirable for the circuit's function and kill it over long time.<sup>3</sup> The balance of intracellular carbon and energy resources in wild-type yeast cells has been optimized by natural evolution.<sup>3</sup> However, capturing cell resources for heterologous pathway gene expression, plasmid maintenance, and product synthesis may interfere with critical cellular processes and result in energetic inefficiency inside the cell, negatively influencing yield of the desired product.<sup>30</sup> To design strains for successful industrial application approaches such as metabolic balancing, respiration improvement, decoupling cell growth and product production phases, and co-utilization of nutrient resources can help to find and resolve metabolic imbalances.<sup>28, 31-32</sup> Therefore, global regulatory networks, metabolic burden, and product toxicity must be considered to achieve high levels of production.<sup>30</sup> However, frequently the connections between these factors involve multiple genes and are not well understood. Therefore, in the next sections

the available tools, their limitations, and the possible solutions to overcome the difficulties discussed above will be highlighted.

### 3.5 Plant Transcription Factors and Promoter Pairs

Regulation of gene expression using TFs impacts many of the biological processes in a cell, such as the cell cycle, metabolic balance, and responses to the environment. Controlling gene expression using TFs is based on the recognition of a promoter sequence by a DNA binding domain (DBD) that is part of the TF.<sup>33</sup> Moreover, proteins without a DBD, which interact with a DNA-binding protein to form a transcriptional complex, are often categorized as TFs. In the last few years, a huge number of (putative) TFs have been identified in plants. These TFs usually form families of proteins that are structurally related.<sup>34</sup> Moreover, they have similar DNA-binding specificities and they are occasionally involved in related phenomena. In some families, formation of homo/heterodimers between TFs results in increasing the variability of the target sequences. The activities of the TFs appear to be modulated by post-translational modification, such as DNA binding, nuclear transport, and proteins-protein interaction. Transcription activators have an activation domain (AD) and DBD, although it is also known that the same TF can act as a repressor and as an activator, depending on, for example, their nuclear concentration in the nucleus or the interacting partners. In common with other eukaryotes, plant TFs (such as DREBs, ARFs and GBF1) that contain domains rich in the acidic amino acids glutamine or proline are transcriptional activators.<sup>33, 35-36</sup> Moreover, the AHA motif (present in heat shock TFs) contains acidic amino acid residues embedding the large hydrophobic amino region, acts as an AD.<sup>37</sup>

### 3.6 Artificial Transcription Factors and Promoter Pairs

Transcriptional regulation is a very important control point for pathway gene expression.<sup>38</sup> Increasingly complex synthesis tasks for engineered organisms needs a wider range of transcriptional activation capacities. Moreover, a distinct promoter is needed to control the expression of each gene in the engineered pathway in eukaryotic cells.<sup>39</sup> That means synbio requires a large collection of promoters, while small set of chemically-inducible native yeast promoters (*e. g. MET3, MET25, PHO5, CUP1, GAL1* and *GAL10* promoters)<sup>40</sup> and yeast constitutive promoter (*e.g. ADH1, TEF1, TEF2, and GPD* promoters)<sup>41</sup> are components of well-established synbio systems in yeast. Furthermore, the efficacy of these systems often suffers from their impact on endogenous yeast regulatory networks.<sup>42</sup> One particularly attractive focus for synbio is developing tools that modify protein levels in response to known input signals,

providing the ability to reprogram the behavior of cell. In nature, TF commonly contains the DBD and AD and upon binding DBDs to a promoter containing the binding site, the downstream gene is expressed.<sup>43</sup> Several efforts have focused on creating a collection of ATFPs to expand the catalogue of available transcriptional controlling tools.<sup>42, 44-45</sup> ATFPs are modular units that facilitate the performance of complex and combinatorial transcriptional regulation. Therefore, the increasing collection of ATFPs provides superior flexibility and control of biological systems. ATFPs contain (i) programmable DBDs (*i.e.* TALE-,<sup>46</sup> CRISPR/ dCas9-derived TFs<sup>45</sup>) or DBDs of natural TFs (ii) ADs and (iii) synthetic promoter containing the binding site of DBD fused to a minimal promoter.<sup>17</sup> However TALE- and CRISPR/ dCas9-based ATF are significantly larger and complex.<sup>42, 45</sup> Additionally, transcriptional activation by CRISPR/Cas9-derived ATFPs is lower than TALE-derived ATFPs.<sup>42</sup> Plant-derived ATFPs are a promising alternative due to their relatively small size and ease of construction. Plant-derived ATFPs, unlike TALE-CRISPR/dCas9-derived ATFPs, cannot be designed to target any desired binding site, because plant DBD needs to be paired with their own binding site. To generate ATFPs using heterologous TFs derived from plants or other organisms in the yeast *S. cerevisiae*, a platform was developed in this study, recently published by ACS synthetic biology journal.<sup>17</sup>

Natural and synthetic biological genetic tools, such as ATFPs, generally obey several design rules.<sup>17, 34, 43</sup> The parts need to be assembled together in specific orders, orientations, and spacing to obtain needed functionalities.<sup>17, 43</sup> These structural requirements can be considered as grammatical rules.<sup>43</sup> For example, *E. coli* promoters typically need RNA polymerase binding boxes at 10- and 35-bp upstream of the transcription start site to start transcription, whereas translation of proteins requires a start codon in the mRNA transcript.<sup>43</sup> ATFPs also need to be designed based on the same grammatical rules. Purcell *et al.* developed a grammar for the design of ATFPs in eukaryotic cells and implemented it within GenoCAD, a Computer-Aided Design (CAD) software for synthetic biology.<sup>43</sup> Their grammar covers the design of ATFs containing effector domains to activate or repress target gene transcription and fluorescent reporter domains to quantify ATF abundance, and it describes the design of ATFs that form dimeric complexes with other ATF(P)s, and thereby cooperativity effects between ATF(P)s can be used to create logical circuits.<sup>43</sup>

### 3.7 Episomal versus Integration

Due to the simplicity of manipulation in the plasmid level, plasmid-based systems are commonly used to introduce heterologous metabolic pathways into *S. cerevisiae*. Specifically,  $2\mu$  plasmids

are preferred due to their higher copy number in comparison to CEN/ARS-based plasmids.<sup>17, 47</sup> Nevertheless, there are major issues related to plasmids, including unpredictability in plasmid copy number between cells and segregational instability, making it difficult to maintain uniform and stable cell populations.<sup>30, 47</sup> Therefore, subpopulations of engineered cells often exist that express the final product at very different levels.<sup>48</sup> Moreover, plasmid-based systems require selectable markers and media that can be too expensive for industrial applications.<sup>48-49</sup> Although a metabolic pathway introduced by a plasmid can be integrated into the host genome, making such a plasmid is laborious in comparison to the recently developed CRISPR/Cas9 integration system that allows multi-fragment integration into a single locus or into multiple independent loci in the yeast genome. Moreover, it is frequently acknowledge that genome integration of a metabolic pathway leads to higher level of production than plasmid integration.<sup>17, 49</sup>

### **3.8 CRISPR/Cas-based Technology: Synbio Gemstone**

*S. cerevisiae* is a suitable host for genome engineering because of its capacity for efficient homologous recombination (HR).<sup>50</sup> Although a number of genome engineering methods for *S. cerevisiae* have existed for years, strain development is still tedious because it requires diverse multidisciplinary techniques. The endogenous HR pathway can use DNA fragments whose ends have homology to genomic sites flanking a double-strand break to replace the endogenous sequence with altered versions.<sup>51</sup> For metabolic engineering projects, fairly extensive multi-step metabolic engineering are typically needed and the rate of endogenous HR achievable in yeast has not been high enough to generate large numbers of modifications.<sup>51</sup> However, construction time could be greatly reduced by establishing techniques that allow high efficiency integration of complex metabolic pathways.

The CRISPR/Cas system is an endogenous component of the prokaryotic immune system that confers resistance to foreign genetic elements.<sup>52</sup> The type II bacterial CRISPR/Cas system implementing Cas9 protein has recently emerged as a technique allowing efficient, simple, and highly specific gene-targeting.<sup>53</sup> In this system, RNA harboring trans-encoded RNA (tracrRNA) and the synthetic guide RNA (gRNA) helps CRISPR-associated protein (Cas9) to recognize and cut exogenous DNA.<sup>52</sup> By delivering the Cas9 nuclease complexed with a synthetic gRNA(s) into a cell, DSB(s) can be introduced in the genome at a desired location(s) allowing existing gene(s) to be removed and/or new one(s) to be added.<sup>53-54</sup> Thereby, Cas9 makes it possible to introduce (knock in) multiple genes in each round of transformation.<sup>55</sup> Cas9 can dramatically stimulate HR-based genetic engineering at specific loci, and therefore genetic engineering with Cas9 is more efficient in comparison to genetic engineering using only native yeast HR.<sup>50, 56</sup> This is quite an

## INTRODUCTION

---

advantage since it removes the need for recycling of selection markers through genome integration steps. Furthermore, industrial producer strains do not contain marker genes, and hence, genetic backgrounds closer to final production strains can be engineered.<sup>50</sup>

CRISPR/Cas9 system have been used as a powerful tool for multi-locus gene knock-out in yeast, where CRISPR/Cas9 system facilitate the integration of multi double-stranded oligonucleotides (dsOligo) containing homology arms to the target sites and an stop codon or frame-shift deletion.<sup>56-57</sup> For example, the HI-CRISPR strategy allows one-step, simultaneous, multiple gene disruptions in *S. cerevisiae*.<sup>56</sup> The HI-CRISPR/Cas9 system contains iCas9 (a variant of wild-type Cas9), tracrRNA, and crRNA array. In the HI-CRIPR system, the crRNA array contains one promoter driving expression of a set of gRNAs flanked by 50-bp left- and right-homology regions (called spacers).<sup>56</sup>

CRISPR/Cas9 can also be implemented to generate transcriptional regulators.<sup>58</sup> The endonuclease inactive (deficient) variant of Cas9 (dCas9) fused to a regulatory domain can be targeted to synthetic promoters through a gRNA and up- or downregulate target genes.<sup>59</sup> For example, if dCas9 is targeted to a promoter of an ORF, then the gRNAs are called as interference gRNAs,<sup>44</sup> or if dCas9 is fused to a repressor domain, then it can behave like a repressor.<sup>44</sup> Alternatively, if dCas9 is fused to an activation domain like VP64, it acts as an ATFP.<sup>51</sup> Recently, functional capabilities of the CRISPR system have expanded further with the development of CRISPR-related nucleases with different gRNA binding and endonucleolytic properties<sup>60</sup> that can function as alternatives to Cas9 in synbio.

Metabolic engineering projects often need overexpression, knock-down, and knock-out of multiple gene targets.<sup>61</sup> As describe above, CRISPR system has been implemented for approximately all the functions needed for metabolic engineering in yeast. Hence, combining several CRISPR-derived functions in the same cell provides synthetic biologists with valuable toolkit for genetic manipulation of multiple targets. For example, Vanegas *et al.* combined Cas9-mediated genome editing and dCas9-derived ATF, allowing yeast to switch between a genetic engineering and a pathway control states (switch on/off system).<sup>62</sup> Moreover, Deaner and Alper established a fast method for fine-tuned expression of enzymes of a metabolic pathway through dCas9-VPR or dCas9-MXI1 regulation (graded modulation of gene expression levels).<sup>63</sup> Deaner *et al.* reprogrammed the dCas9-VPR activator to simultaneously activate and repress multiple targets within a yeast cell (dual-mode activator/repressor).<sup>64</sup>



Recently, Lian *et al.* developed an orthogonal tri-functional CRISPR system, called CRISPR-AID, based on one nuclease-deficient CRISPR protein fused with an AD for transcriptional activation (CRISPRa), a second nuclease-deficient CRISPR protein fused with a repression domain for transcriptional interference (CRISPRi), and a third catalytically active CRISPR protein for gene deletion (CRISPRd).<sup>65</sup>

Despite being relatively new technology, both the genetic engineering and gene regulatory capabilities of CRISPR/Cas9 have already proven to be extremely valuable in metabolic engineering projects.<sup>44, 59, 66</sup>

### **3.9 Metabolic Engineering: White Biotechnology**

The field of metabolic engineering implements genetic tools and circuits to manipulate microbial metabolism to produce compounds of interest.<sup>67</sup> Therefore metabolic engineering is promising areas of research that facilitated the expansion of industrial biotechnology. Recent developments in synbio provides novel metabolic engineering toolbox for pathway design and provides researchers and companies with implementing yeast for producing various chemicals.<sup>18</sup> The earliest example of metabolic engineering dates back to 1973, when Cohen and Boyer successfully introduced heterologous genes into a bacterial cell, and it became obvious that microorganisms could be used as little chemical factories for cost-effective production of many products.<sup>68,69</sup> Recent dramatic progresses in industrial biotechnology and numerous successful applications of new industrial procedures have resulted in important progress of the field of so-called white biotechnology.<sup>70</sup> Nowadays, by rising oil costs and fears about climate change, it is necessary to think about greener sources for transportation fuels.<sup>71</sup> Furthermore, many foods, pharmaceuticals, and cosmetic ingredients are extracted from plants, and therefore we are dependent on seasonal growth, and natural resources are prone to depletion which has brought about worldwide concern.<sup>72</sup> Moreover, physical and chemical extraction methods can be expensive.<sup>73</sup> Many companies, both biotech and traditional chemical companies, are now translating research successes from academic and industrial groups into industrial processes. They are interested in producing chemicals that are environmentally friendly, less expensive, and have superior properties compared with those produced by nature or chemically synthesized.<sup>61</sup> Examples include Sanofi Aventis (artemisinic acid used as an anti-malarial drug),<sup>8</sup> Gevo, Butamax (isobutanolis used as biofuel),<sup>74</sup> Evonik, ADM, CJ, Ajinomoto (lysine used as feed additive).<sup>61</sup>

### 3.10 Carotenoids

Carotenoids are a subgroup of isoprenoid compounds comprising over 700 structures, including  $\beta$ -carotene.<sup>75-77</sup> In nature, carotenoids are synthesized by phototrophic organisms, but also by many non-phototrophic species (except animals).<sup>77</sup> All chlorophototrophs, including several eubacterial phyla, algae and plants, employ bacteriochlorophyll and/or chlorophyll for light harvesting to produce carotenoids at high level. In animals, carotenoids are used to produce many of the bright colours. These organisms, including humans, have to obtain them from their diet, as they cannot synthesize carotenoids.<sup>76</sup>

The mevalonate pathway<sup>78</sup> uses three acetyl-CoAs to form hydroxyl-methyl-glutaryl-CoA (HMG-CoA) that is reduced to mevalonate by the action of HMG-CoA reductase (HMGR), a rate-limiting enzyme.<sup>79</sup> Consecutive phosphorylations and decarboxylation of the mevalonate result in isopentenyl pyrophosphate (IPP) and its isomer, dimethylallyl pyrophosphate (DMAPP). IPP and DMAPP are further condensed by prenyl transferases to synthesize geranyl diphosphate (GPP), FPP, and GGPP. A variety of terpene synthases use GPP, FPP, and GGPP to synthesize monoterpenes, sesquiterpenes, and diterpenes including carotenoids, respectively.<sup>80-81</sup> Carotenoids can function as accessory pigments in photosynthesis, augmenting light harvesting. More specifically,  $\beta$ -carotene whose pigmented products are observable in yeast colonies, making them tractable model products for testing combinatorial approaches, used for the case study here. Furthermore, it is also the precursor of vitamin A and can be converted into many other secondary metabolites, including  $\beta$ -ionone, astaxanthin, and canthaxanthin, zeaxanthin, and safranal.<sup>81</sup>  $\beta$ -Ionone is an aroma apo-carotenoid used in flavors.<sup>82</sup> Moreover, carotenoids are used as food-coloring agents, antioxidants, in aquacultures, in cosmetics, and in the pharmaceutical industry.<sup>41</sup>

### 3.11 Naringenin

Naringenin (NG) is key intermediate flavonoid.<sup>83</sup> The flavonoid group of secondary metabolites formed from phenylpropanoid and fatty acid derivatives. Flavonoids are UV-B protectors,<sup>84</sup> attractors or deterrents of insects for or from feeding, attractors of pollinators, signal molecules in plant-bacteria interactions<sup>85</sup> and are used for human pathologies treatment.<sup>86</sup>

In plants, flavonoids are derived from L-tyrosine and L-phenylalanine. Para (p)-coumaric acid is formed from L-tyrosine. The tyrosine ammonia-lyase (TAL) and 4-coumaroyl-CoA ligase (4CL) convert L-tyrosine into p-coumaroyl-coenzyme A (CoA). The biosynthetic pathway is directed to

NG by chalcone synthase (CHS) converting one molecule of p-coumaroyl-CoA and three molecules of malonyl-CoA (synthesized from acetylCoA by a carboxylase) into naringenin chalcone. Next, chalcone isomerase (CHI) isomerizes naringenin chalcone into NG. Additionally, CHS and chalcone reductase (CHR) converts one molecule of p-coumaroyl-CoA and three molecules of malonyl-CoA into isoliquiritigenin.<sup>83</sup>

However, inadequacies in chemical synthesis and extraction from natural sources limited availability of pharmaceutical flavonoids. Therefore, demand for the microbial production of flavonoids has been prominent as a probable alternative approach for flavonoid production.<sup>87</sup> Till now, NG pathway has been engineered in both *E. coli* or *S. cerevisiae*.<sup>87-88</sup> Recently, one study reported the synergistic co-culture system between *E. coli* and *S. cerevisiae* for the production of NG.<sup>89</sup>

### 3.12 Rational Techniques for Metabolic Engineering: Classical Methods

So-called rational approaches for metabolic engineering involve iterative cycles of synthesis and analysis, in which increasingly refined strains are designed and constructed.<sup>90</sup> These approaches are based on prior knowledge about metabolic network topographies and its regulation, bottlenecks and competitive reactions.<sup>90</sup> Till now, most metabolic engineering projects have employed these kinds of rational approaches, and there are a number of chemicals produced on a commercial level using classical metabolic engineering techniques.<sup>91-92</sup> For example, from the area of biofuels is the engineering of *S. cerevisiae* to convert xylose, a carbohydrate that is not used in nature by yeast, into biomass or ethanol.<sup>93</sup>

### 3.13 Non-rational Techniques for Metabolic Engineering

Generating an optimal microbial cell factory usually requires overexpression, downregulation, and deletion of multiple genes.<sup>90</sup> When done sequentially, such rewiring of cellular metabolism with classical methods is time-consuming and low throughput and it is possibly optimized pathways in laboratory strains cannot be easily transferred to industrial strains, as laboratory and industrial strains typically have different genetic backgrounds.<sup>94</sup> Where naturally occurring alternatives are not known, non-rational methods can be used to bridge biological knowledge gaps.<sup>95 96</sup>

Non-rational methods are high-throughput techniques that do not depend on *a priori* knowledge about the optimal amount of intermediates for the engineered pathway.<sup>97</sup> For example, random or semi-random genetic changes are induced, and the modified strains are screened for the

desired properties. Later, these techniques were widely used for industrial strains. Using this approach, penicillin production in *Penicillium chrysogenum* was increased to 50 g/liter, 4,000-fold more than that of the original strain.<sup>98</sup> Some non-relational methods also allow testing of many genetic perturbation combinations simultaneously, so-called combinatorial approaches. The most powerful of these allow modification of the host genome in a modular, parallel, and high-throughput manner. Till now, several combinatorial metabolic engineering strategies have been reported that allow pathway optimization, for example, at the DNA, RNA and protein levels by modifying the factors influencing duplication, transcription and translation procedures.<sup>65, 99-103</sup>

Inverse metabolic engineering is an extension of combinatorial optimization that allows the separation and identification of the mutations linked to the desired phenotype. This information facilitates rebuilding of the strain via classical methods by providing detailed knowledge about the metabolic network.<sup>104-105</sup> Three key factors were pointed that affect the success of inverse metabolic engineering: (i) a high-throughput monitoring method for the desired phenotype; (ii) a library with sufficient genetic diversity to greatly increase the chance of detecting the phenotype of interest; and (iii) a rapid method to detect those genetic modifications that are responsible for the expression of the desired phenotype. Advancements in microarray technologies and DNA sequencing have helped facilitate the last factor.<sup>106-107</sup> However, the first factor is often much more difficult to achieve because the phenotype of interest is often a small molecule, and monitoring a library of thousands of mutants using traditional analysis methods like HPLC and GC-MS is laborious. Therefore, establishment of high-throughput screening methods is important.<sup>108</sup>

### **3.14 Biosensor: Cell Factory Lighter**

The development of microbial strains able to produce interesting metabolites at sufficient levels has been facilitated by combinatorial approaches.<sup>32, 109-110</sup> However, within the Design-Build-Test Cycle for strain engineering, selection of strains with the highest product yield from a library with huge genetic variety is still a serious bottleneck. Many synthetic biologists are hoping that progress in the scope of metabolite sensing can introduce high throughput methods to remove the barriers caused by slow technologies (*i.e.* requiring chromatography). Biosensors may become a larger part of the solution. Biosensors are novel genetically encoded circuits that detect metabolites in an *in vitro* or *in vivo* manner and translate their concentrations into more quantifiable signals.<sup>108</sup> For example, increased production of an intracellular metabolite can be sensed by a biosensor-TF.<sup>111</sup> In one recently published work, prokaryotic LTTR TFs were used

as biosensors to show CCM and NG production level in other prokaryotes.<sup>112</sup> Upon interaction between LTTR TFs and a synthetic promoter with corresponding BS; its downstream fluorescent proteins was expressed. Subsequently, FACS was used to screen fluorescent output.<sup>112</sup> In 2016, Skjoedt *et al.* used the same LTTR TFs to show CCM or NG productions in the yeast *S. cerevisiae* strains engineered with their biosynthetic pathways.<sup>113</sup>

Current attempts in the area of biosensor development for metabolic engineering are extremely concentrated on TF-based detection. Riboswitches - RNA-based 'sensors'- are another class of biosensor. In nature, external signals enhance structural alterations in the 5'- untranslated region of mRNA leading to differential gene expression.<sup>21</sup> Synthetic riboswitches have been developed that respond to diverse compounds.<sup>114-115</sup> Coupling a specific riboswitch with a reporter gene system allows the amount of product from a biosynthetic pathway to be decoded into a measureable reporter output.<sup>115</sup>

In recent years, stains, dyes and chromogenic substrates have been successfully used as biosensors, but accessibility of chemistries and biological compatibilities is often finite their application.<sup>116</sup> However, these orthogonal substrates can be combined with FACS-based detection and sorting and thereby providing a promising approach to distinguish metabolites and enzymes in high throughput manner.

### **3.15 A Vision of the Future of Synbio**

In the first wave of synbio, basic tools were combined to form small circuits with specified behaviors including transcription, translation and post-translational processes. Many of the genetic circuits described thus far in the field have been simple and are usually aimed at controlling isolated cellular functions.<sup>16, 18</sup> Now the second wave of synbio is emerging, in which basic parts and tools are being integrated to create systems-level circuitry.<sup>19</sup>

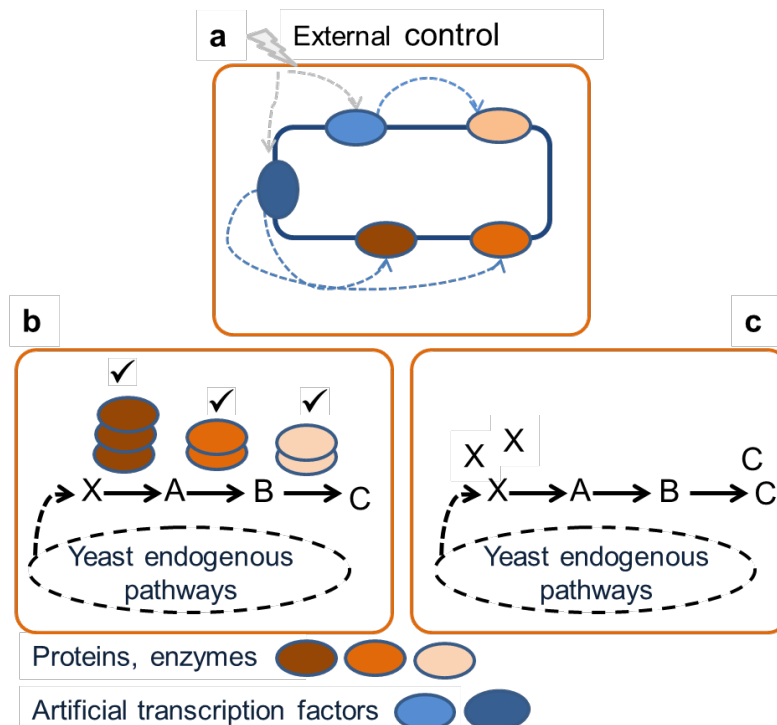
Friedland *et al.* (2009) described logical circuits able to program cells by means of modern computing principles.<sup>117</sup> Recently researchers from Harvard Medical School and the Wyss Institute for Biology developed the first molecular recorder using the CRISPR system that allows storage of massive amounts of digital information in the genome of bacteria. If cells can store information, the potential applications are vast. For example, bacteria can survive in diverse environments, and therefore programmed microorganisms could be spread on soil to record environmental pollution and neurons could be used to record brain development in a living organism.<sup>9</sup>

However, despite the huge growth in synbio, major challenges still exist in the design-build-test cycle.<sup>54</sup> These difficulties principally stem from the lack of predictive models because of the nonlinearity of biological systems, restricted libraries of parts, and low-throughput characterization methods. Furthermore, it is not always clear how to control noise and how to transfer the functionality of engineered elements between organisms. We need new design computational tools that use biological variability, uncertainty, and evolution to develop synthetic systems that are more reliable. If we could address all of these issues, synbio would be able to have an important and fundamental impact on our ability to solve biological and environmental problems.

### 3.16 Goals of this Study

Synbio projects often need variable expression of various genes to create and improve biosynthetic pathways or construct multi-subunit cellular complexes. Hence, three novel tools and approaches were developed here. The first of these is a set of inducible, plant-derived ATFPs for orthogonal transcriptional regulatory systems (**Figure 1a**, Chapter 4). More specifically, a set of 106 synthetic transcription regulatory units based on plant TFs and their cognate binding sites was generated, covering a wide range of transcriptional outputs. Some transcriptional control units confer expression levels which are 6- to 10-fold stronger than that of the widely used strong yeast *TDH3* promoter. The diverse regulatory units were assembled from DBDs or full-length coding sequences of plant TFs, fused to yeast, virus and novel plant-based ADs, and one, two and four copies of their binding sites. The second novel tool is a method for combinatorial optimization of heterologous pathways without *a priori* knowledge of the optimal expression levels of all genes in the synthetic gene regulatory network. Named COMPASS (**Figure 1b**, Chapter 5), it is a high-throughput cloning method for the empirical balancing of metabolic pathway gene expression in the yeast *S. cerevisiae*. The utility of COMPASS combinatorial cloning approach was shown with co-expression of a four- and five-gene pathway to generate yeast cells producing  $\beta$ -ionone and NG at high levels. Finally, CaPRedit (**Figure 1c**, Chapter 6) was developed as an approach for speeding up the strain modification and it makes possible for very low to very high-expression of wanted enzymes in desired time in yeast through implementing CRIPSR/Cas9-intermediate genome editing for one-step integration of multiple plant-derived ATFPs modules. CaPRedit can be used for redirection of endogenous metabolic flux toward a desired product in the yeast *S. cerevisiae*. The FPP production was optimized. FPP is a central precursor of nearly all yeast isoprenoids,<sup>81</sup> and is used to produce diverse products, e.g. alcohols, hydrocarbons. Isoprenoids is big family of natural products involved in

variety of important biochemical functions processes including electron transport, maintaining membrane fluidity, subcellular protein targeting and regulation, and cellular development, sterol and. The plant-derived ATFPs were employed to control expression of genes in the synthetic pathway, but other regulators such as TALE- and CRISPR/Cas9-derived ATFPs can be easily implemented as well for COMPASS and CaPRedit methods.



**Figure 1.** Tools developed in this study. **(a)** Plant-derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces cerevisiae*. **(b)** Rapid combinatorial optimization of biochemical pathways based on artificial transcription factors. **(c)** CaPRedit: Genome editing using CRISPR-Cas9 and plant-derived transcriptional regulators for the redirection of flux through the FPP branch-point in yeast

#### 4 Plant-derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces cerevisiae*

Gita Naseri<sup>1</sup>, Salma Balazadeh<sup>2,3</sup>, Fabian Machens<sup>1</sup>, Iman Kamranfar<sup>3</sup>, Katrin Messerschmidt<sup>1</sup>, Bernd Mueller-Roeber<sup>2,3,\*</sup>

<sup>1</sup>University of Potsdam, Cell2Fab Research Unit, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany; <sup>2</sup>Max-Planck Institute of Molecular Plant Physiology, Plant Signalling Group, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany; <sup>3</sup>University of Potsdam, Department Molecular Biology, Karl-Liebknecht-Str. 24-25, House 20, 14476 Potsdam, Germany

\*Corresponding author: Bernd Mueller-Roeber, bmr@uni-potsdam.de

Running title: Plant transcription factors for yeast synthetic biology

Key words: *Arabidopsis thaliana*, artificial transcription factor, NAC transcription factor, synthetic biology, plant

**The manuscript has been published in ACS Synthetic Biology:**

Naseri, G. *et al.* (2017). Plant-Derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces cerevisiae*. *ACS Synth Biol.* 6, 1742-1756.



# Plant-Derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces cerevisiae*

Gita Naseri,<sup>†</sup> Salma Balazadeh,<sup>‡,§</sup> Fabian Machens,<sup>†</sup> Iman Kamranfar,<sup>§</sup> Katrin Messerschmidt,<sup>†</sup> and Bernd Mueller-Roeber<sup>\*,‡,§</sup>

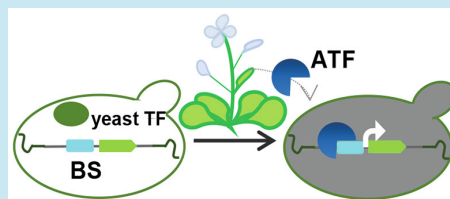
<sup>†</sup>Cell2Fab Research Unit, and <sup>§</sup>Department Molecular Biology, University of Potsdam, Karl-Liebknecht-Straße 24-25, Potsdam, 14476, Germany

<sup>‡</sup>Plant Signalling Group, Max-Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, Potsdam, 14476, Germany

## Supporting Information

**ABSTRACT:** Control of gene expression by transcription factors (TFs) is central in many synthetic biology projects for which a tailored expression of one or multiple genes is often needed. As TFs from evolutionary distant organisms are unlikely to affect gene expression in a host of choice, they represent excellent candidates for establishing orthogonal control systems. To establish orthogonal regulators for use in yeast (*Saccharomyces cerevisiae*), we chose TFs from the plant *Arabidopsis thaliana*. We established a library of 106 different combinations of chromosomally integrated TFs, activation domains (yeast GAL4 AD, herpes simplex virus VP64, and plant EDLL) and synthetic promoters harboring cognate *cis*-regulatory motifs driving a *yEGFP* reporter. Transcriptional output of the different driver/reporter combinations varied over a wide spectrum, with EDLL being a considerably stronger transcription activation domain in yeast than the GAL4 activation domain, in particular when fused to *Arabidopsis* NAC TFs. Notably, the strength of several NAC–EDLL fusions exceeded that of the strong yeast *TDH3* promoter by 6- to 10-fold. We furthermore show that plant TFs can be used to build regulatory systems encoded by centromeric or episomal plasmids. Our library of TF–DNA binding site combinations offers an excellent tool for diverse synthetic biology applications in yeast.

**KEYWORDS:** *Arabidopsis thaliana*, artificial transcription factor, NAC transcription factor, synthetic biology, plant



The eukaryotic yeast *Saccharomyces cerevisiae* with its relatively small and well-characterized genome (12 Mb) is particularly well suited for biological engineering.<sup>1</sup> It is often employed as a biofactory for the production of enzymes, biofuels, or pharmaceutical and nutraceutical ingredients.<sup>2</sup> For metabolic engineering purposes, efficient and diverse cellular regulation systems are required,<sup>3,4</sup> in particular to avoid the constitutive expression of pathway enzymes that may present a metabolic burden on the cell.<sup>1</sup> Clearly, diverse regulatory mechanisms for the control of (heterologous) gene expression are crucial for pathway engineering or the creation of complex protein expression systems in synthetic biology.<sup>3–5</sup>

Transcriptional regulation plays an important role in gene expression,<sup>3,6</sup> and therefore the establishment of orthogonal regulation systems based on artificial transcription factors (ATFs) for use in synthetic biology applications is crucial. Transcription factors (TFs) typically contain at least two functional domains, namely the DNA-binding domain (DBD) that binds to the promoters of target genes, and the activation domain (AD) that activates transcription by interacting with the basal transcription machinery of the cell.<sup>7</sup> TFs bind to specific *cis*-regulatory elements residing in the promoters of target genes to control their spatial and temporal patterns of expression.<sup>5,6,8</sup> While natural TFs control the growth, development, and function of an organism,<sup>8</sup> ATFs are in most cases employed for the coordinated expression of transgenes needed

for metabolic engineering and the construction of synthetic regulatory circuits.<sup>4</sup> ATFs are constructed by joining different DBDs with different ADs and, in eukaryotic cells, a nuclear-localization signal (NLS) that targets the chimeric regulator to the nucleus.<sup>5</sup> For specific DNA-binding, DBDs of TFs with known *cis*-regulatory motifs or programmable DBDs designed to target predefined binding sites can be employed.<sup>5–8</sup> For example, transcription activator-like effectors (TALEs) and the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/dead CRISPR-associated 9 (dCas9) systems are now employed to generate synthetic TFs.<sup>6</sup> Generating TALE-based ATFs is, however, labor-intensive because of the TALEs' repeat structure and their relatively large size (>100 kDa).<sup>9</sup> Although the simplicity of small-guide RNA (sgRNA) design makes the CRISPR/Cas9 system a powerful tool for genomics research, CRISPR-derived ATFs in general lead to a lower degree of transcriptional activation than TALE-derived ATFs<sup>6</sup> (and own observations). In addition, off-target effects are an important consideration for both TALE and dCas9 transcriptional platforms.<sup>6</sup>

For the regulation of complex gene expression systems in synthetic biology projects, ATFs with different capacities for

Received: March 24, 2017

Published: May 22, 2017

transcriptional activation are needed. Moreover, as eukaryotic organisms generally do not employ polycistronic mRNAs a distinct promoter is required to control the expression of each gene in the host organism (e.g., yeast).<sup>10</sup> Thus, developing orthogonal ATFs on the basis of heterologous (nonyeast) TFs, such as those derived from plants, is an important goal.

Plant TFs are grouped into diverse families according to conserved motifs that define their DBD. More than 2000 TFs belonging to more than 60 different families are known in higher plants (e.g., <http://plntfdb.bio.uni-potsdam.de/v3.0>), and nearly half of them are considered plant-specific. Many plant TF family members characterized to date control genes involved in cellular differentiation, organ development, and the response to environmental stresses.<sup>8</sup> Considering the wide evolutionary distance between *S. cerevisiae* and plants (with a common ancestor at least one billion years ago), TFs from plants are likely good candidates for the construction of ATFs for use in synthetic biology approaches in yeast; plant-derived ATFs and their binding sites are expected to show minimal interference with endogenous yeast TFs.

Currently, the most widely used ADs for target gene activation in yeast are the herpes simplex virus protein VP16 (or VP64 if containing four VP16 copies) and the yeast GAL4 activation domains.<sup>5,6,11</sup> Previously, it was shown that a 24-amino acid long motif from the *Arabidopsis* ERF98/TDR1 transcription factor, named “EDLL” due to the presence of conserved glutamic acid (E), aspartic acid (D), and leucine (L) residues, acts as a strong AD in plants.<sup>12</sup> However, whether EDLL is functional in yeast has to our knowledge not yet been reported.

In *S. cerevisiae*, the transactivation capacity of ATFs is typically characterized after integration of the expression control circuits into the nuclear genome;<sup>5</sup> within the circuit, modified *GAL1* promoters are often used to control the expression of ATFs which then control the expression of their target promoters.<sup>5,13,14</sup> While constitutive expression of TFs and pathway genes may put a metabolic burden on the cell and thereby negatively affect growth,<sup>1</sup> nutrient-/small molecule-responsive DNA sequences allow conditional induction of *ATF* genes.<sup>5,13,14</sup> In addition to, for example, isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG)-responsive transcriptional units, a regulatory module responsive to anhydrotetracycline (ATc) has been developed previously.<sup>13,14</sup> For the heterologous expression of multiple genes in yeast, plasmid-dependent systems are often employed because of easy manipulation. However, plasmid-born control units often lack sufficient robustness due to segregational and/or structural instability, adding to cell-to-cell variations stemming from different patterns of segregation of plasmids and the endogenous chromosomes.<sup>15</sup> Another aspect of consideration is that the integration of transgenes into the chromosomal DNA often leads to more product than an integration into plasmids.<sup>15,16</sup>

In the present study, we sought to generate a new class of ATFs for synthetic biology projects in yeast, using plant TFs. Moreover, we report that the plant EDLL motif can be successfully employed as a strong AD in yeast. We generated a library of plant-derived ATFs and promoter pairs using different families of plant TFs. Individual ATFs together with their cognate promoters were characterized by integrating transcriptional circuits into the *S. cerevisiae* genome. A wide range of transcriptional outputs was obtained for different plant-derived ATFs. Furthermore, we characterized the expression of two ATFs encoded by either centromeric or

episomal plasmids. We show, that within the circuit, IPTG- and ATc-inducible *GAL1* promoters can be used for the independent transcriptional activation of two plant-derived ATFs.

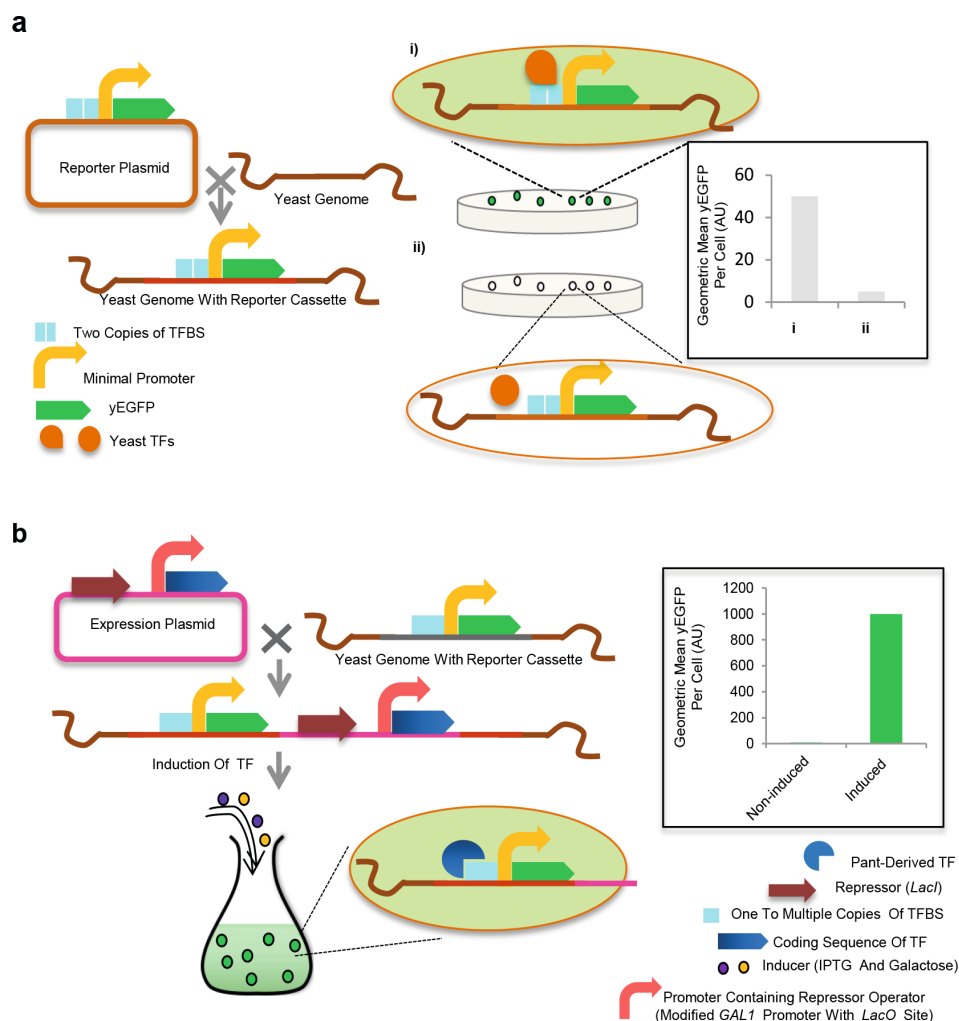
## RESULTS

Regulation at the transcriptional level is a critical step for establishing heterologous metabolic pathways in target organisms.<sup>3,5</sup> Despite some success in the past regarding proteins designed to target synthetic DNA sequences,<sup>5,6</sup> engineering new ATFs with high activity and specificity is still a major concern for synthetic biologists. Here, we attempted to utilize the capacity of heterologous TFs from plants to build ATFs for the transcriptional activation of target genes in yeast (*Saccharomyces cerevisiae*). To this end, we first tested plant *cis*-regulatory elements for background activity in yeast and selected those showing low activity in the absence of the plant TF.

We then established a set of 106 transcriptional control units where plant-derived ATFs are combined in different ways with their cognate binding sites to achieve a wide spectrum of transcriptional outputs with in some cases up to 6- to 10-fold stronger transcription activation than observed for the strong yeast *TDH3* promoter.

**Plant Transcription Factors for Establishing Orthogonal Transcription Units in Yeast.** To construct orthogonal transcription units for use in yeast we employed TFs from the plant *Arabidopsis thaliana*. We selected 14 candidates from different TF families, namely: AINTEGUMENTA (ANT),<sup>17</sup> Related to ABI3/VP1 1 (RAV1),<sup>18</sup> Dehydration-Responsive Element-Binding Protein 2A (DREB2A),<sup>19</sup> LEAFY (LFY),<sup>20</sup> WRKY6,<sup>21</sup> Growth-Regulating Factor 7 (GRF7)<sup>22</sup> and GRF9,<sup>23</sup> the NAC TFs JUNGBRUNNEN1 (JUB1),<sup>24</sup> ORESARA1 (ORE1),<sup>25</sup> ANAC032,<sup>26</sup> *Arabidopsis thaliana* Activating Factor 1 (ATAF1),<sup>27</sup> and ANAC102,<sup>28</sup> DNA-binding-with-one-finger 1 (DOF1),<sup>29</sup> and MYB domain containing protein 61 (MYB61).<sup>30</sup> AGI codes of the TFs and their *cis*-regulatory target sites (core motifs) are given in [Supplementary Table S1](#).

**Test for Activation of Plant *Cis*-regulatory Elements by Endogenous Yeast TFs.** The core *cis*-regulatory motifs of TFs are generally short (in the range of a few base pairs) and therefore frequently occur throughout a given genome. This makes it difficult to predict whether TFs from yeast will bind to the heterologous (plant) TF binding sites and activate expression of a gene linked to it. We therefore established a yEGFP (yeast enhanced green fluorescent protein)-based reporter system to test the effect of plant regulators in the yeast background. On the basis of previous work,<sup>5</sup> we developed a set of plasmids to test the functionality of plant TFs after expression in the microbial host. The plasmids allow the sequential integration of synthetic promoters harboring plant TF binding sites (upstream of the yEGFP reporter) and the coding sequence (CDS) of the corresponding ATFs into the *ura3–52* locus of the yeast genome. Activation of the yEGFP reporter by endogenous yeast TFs (which is undesired) would be detected by background yEGFP fluorescence in the absence of the cognate ATF. In contrast, specific recognition and activation of the *cis*-regulatory motif in the presence of the ATF, representing orthogonality, would lead to high reporter expression. Plant TFs whose binding sites are not targeted (and hence not activated) by endogenous yeast TFs are selected for establishing synthetic expression regulation control systems functional in yeast. In the approach used here, expression of



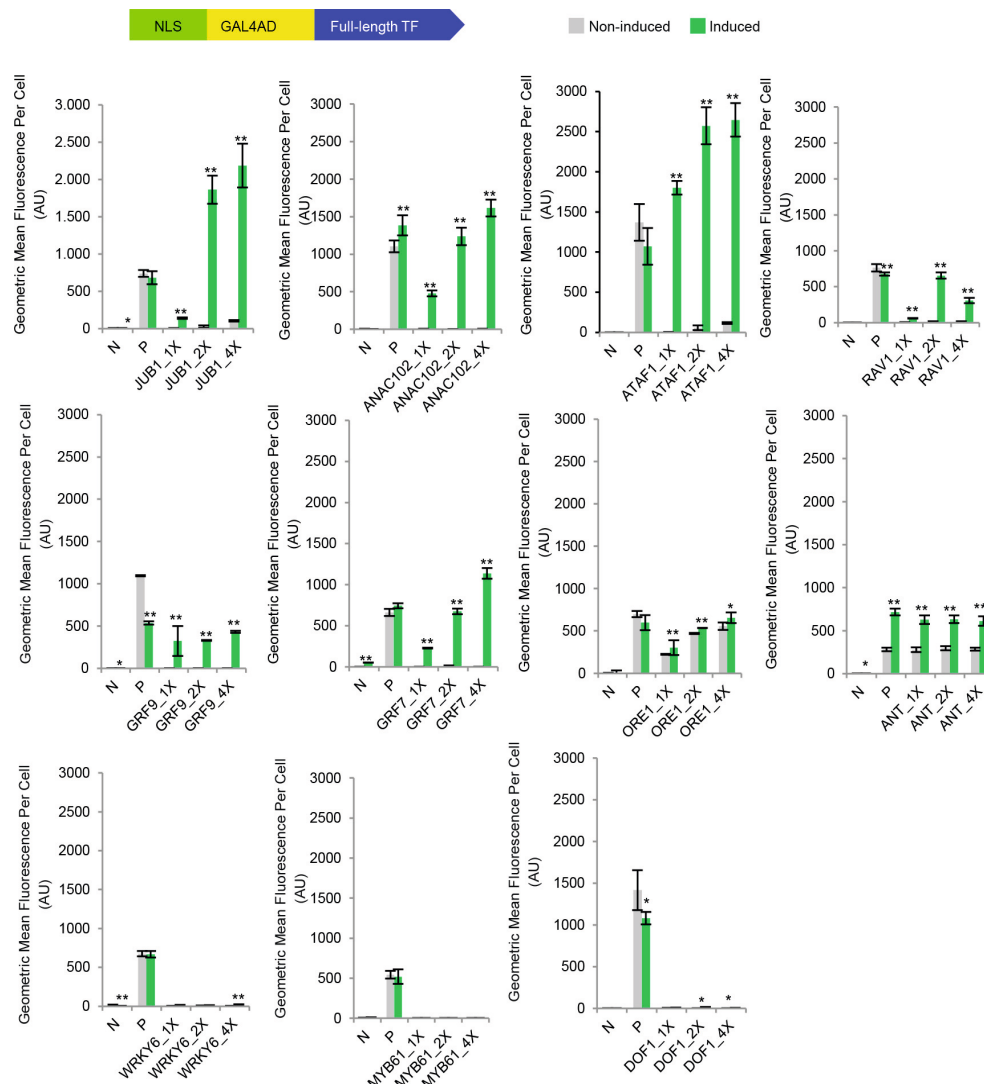
**Figure 1.** Platform for establishing orthogonal transcriptional units using heterologous TFs. (a) Genome integration of yEGFP reporter: A reporter construct harboring two copies of a plant TF binding site (TFBS) upstream of the *CYC1* minimal promoter is integrated into the yeast (*S. cerevisiae*) genome, leaving two possible scenarios: (i) One or more endogenous yeast TFs bind to the heterologous TFBS thereby activating the yEGFP reporter leading to relatively high background expression. (ii) In the absence of yeast TFs binding to the plant TFBS, low reporter expression is observed. Plant TFs, the binding sites of which are not activated by yeast TFs are selected for the construction of artificial TFs (ATFs). (b) Coding sequences of selected plant TFs (entire TF or only DNA-binding domain, DBD) were cloned into the expression plasmid (with or without NLS; with or without activation domain) and placed downstream of the modified *GAL1* promoter containing the *LacO* operator. The plasmids were integrated into the genome of yeast strains harboring the yEGFP reporter cassette with one or multiple copies of the TFBS. Constitutive expression of the repressor (*LacI*) inhibits expression of plant-derived ATFs, while the addition of inducers (IPTG and galactose) results in ATF expression. Binding of the ATF to its cognate BS within the *CYC1* minimal promoter drives yEGFP expression. Fluorescence output is measured in the absence and presence of inducer.

plant-derived ATFs is controlled by an IPTG-inducible *GAL1* promoter.<sup>14</sup>

To assess the activation specificity for binding sites of the 14 plant TFs in yeast, we inserted two copies of their cognate binding sites upstream of the *CYC1* minimal promoter in plasmid pGN005, and integrated the constructs into the chromosomal *ura3–52* locus. Basal yEGFP reporter activity in the absence of plant TFs was determined by flow cytometry (see Figure 1 for the experimental outline). Relatively high basal reporter output was observed for the binding motifs of ANAC032, DREB2A, and LFY indicating activation of reporter gene expression by endogenous yeast TFs. These TFs were therefore excluded from further studies. In contrast, promoters harboring binding sites for RAV1, WRKY6, GRF9, GRF7,

ANAC102, DOF1, MYB61, ANT, ATAF1, JUB1, and ORE1 showed only minimal reporter output (Supplementary Figure S1); the corresponding TFs were therefore chosen for further investigation.

**Transcriptional Competence of Plant-Derived TFs in Yeast.** The 11 TFs identified above were next analyzed for their ability to drive gene expression in yeast. To this end, we inserted the full-length CDS of each TF downstream of a nuclear localization signal (NLS; from the SV40 large T antigen) and the *GAL4* AD, thereby creating NLS-*GAL4*AD-TF fusions (Figure 2). We also inserted all TFs between the NLS and a downstream located *GAL4* AD, thereby creating NLS-TF-*GAL4*AD fusions (Supplementary Figure S2).



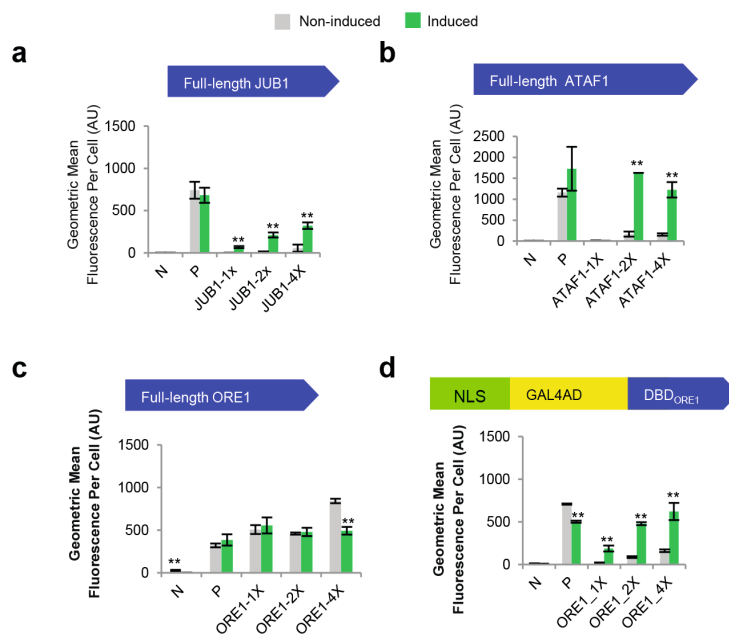
**Figure 2.** Transcriptional output of plant-derived NLS-GAL4AD-TF regulators. The CDSs of plant TFs were inserted downstream of an N-terminal NLS and a GAL4 AD. The NLS-GAL4AD-TF regulators were tested for their capacity to activate *yEGFP* reporter gene expression from the *CYC1* minimal promoter harboring one (1×), two (2×) or four (4×) copies of the cognate binding sites. The effector and reporter constructs were chromosomally integrated into the yeast genome. *yEGFP* output signal was tested in the absence and presence of inducer (IPTG). Gray, noninduction medium; green, induction medium. N, negative control (*Pro<sub>CYC1 min</sub>-yEGFP*); P, positive control (*Pro<sub>TDH3</sub>-yEGFP*). Mean fluorescence intensity per cell is given. Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Asterisks indicate statistically significant difference from noninduction medium (Student's *t*-test; (\*)  $p < 0.05$ ; (\*\*)  $p < 0.01$ ). AU, arbitrary units. Full data are shown in [Supplementary Data S1](#).

*yEGFP* reporter output revealed strong transcriptional activation capacity for several NLS-GAL4AD-TFs, in particular those derived from the NAC factors JUB1, ANAC102, ATAF1, and NLS-GAL4AD-GRF7. In these cases, transactivation capacity generally increased with increasing TF binding site numbers. For NLS-GAL4AD-JUB1 and NLS-GAL4AD-ATAF1 reporter output after IPTG induction considerably exceeded that of the positive control that expresses *yEGFP* from the strong constitutive yeast *TDH3* promoter (*Pro<sub>TDH3</sub>-yEGFP*), in particular when multiple TF binding sites were placed upstream of the *CYC1* minimal promoter. Transactivation strength was less pronounced for NLS-GAL4AD-TFs derived from RAV1, GRF9, or ORE1, and transactivation capacity of NLS-GAL4AD-ANT was independent of the binding site copy

number. Of note, NLS-GAL4AD-TFs derived from WRKY6, MYB61, and DOF1 did not show any transcription activation property (Figure 2).

In the case of NLS-TF-GAL4AD fusions, high transcription activation capacity was again observed for ATFs based on JUB1, ATAF1, and GRF7, but also for GRF9 and ANT, while NLS-ANAC102-GAL4AD was transcription activation silent (Supplementary Figure S2). ATFs derived from RAV1 and ORE1 showed moderate activation capacity, while virtually no transcription activation capacity was detected for NLS-TF-GAL4AD fusions built on WRKY6, MYB61, and DOF1 (Supplementary Figure S2), as for the NLS-GAL4AD-TF versions of these TFs (Figure 2). These three latter TFs were therefore excluded from further studies.





**Figure 3.** Transcriptional output of native plant TFs. The native NAC transcription factors (a) JUB1, (b) ATAF1, and (c) ORE1 were tested for their capacity to activate *yEGFP* reporter gene expression from the *CYC1* minimal promoter harboring one (1X), two (2X) or four (4X) copies of the cognate binding sites. The effector and reporter constructs were chromosomally integrated into the yeast genome. The *yEGFP* output signal was tested in the absence and presence of inducer (IPTG). Gray, noninduction medium; green, induction medium. (d) Transcriptional output of DBD<sub>ORE1</sub> fused to the GAL4 AD. The ability of the DBD<sub>ORE1</sub>-GAL4AD regulator to activate gene expression from one, two, or four ORE1 binding sites was tested in the absence and presence of IPTG. Gray, noninduction medium; green, induction medium. N, negative control (*Pro<sub>CYC1</sub> min<sup>-</sup>yEGFP*); P, positive control (*Pro<sub>TDH3</sub>-yEGFP*). Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Asterisks indicate statistically significant difference from noninduction medium (Student's *t*-test; (\*)  $p < 0.05$ ; (\*\*)  $p < 0.01$ ). AU, arbitrary units. Full data are shown in [Supplementary Data S2](#).

The fact that NLS-GAL4AD-ANAC102 shows strong (Figure 2), but NLS-ANAC102-GAL4AD shows no transactivation capacity (Supplementary Figure S2) highlights the fact that the position of the TF relative to the GAL4 AD (N- versus C-terminal) can have a strong effect on the binding and/or transactivation capacity of plant-derived ATFs.

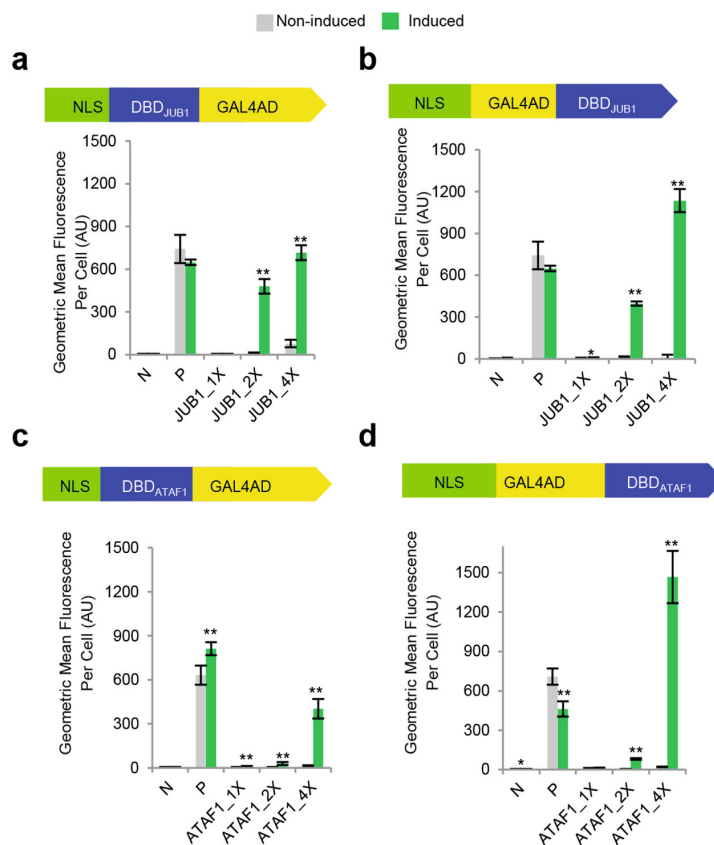
A surprising observation was that both ORE1-derived TFs, that is, NLS-GAL4AD-ORE1 and NLS-ORE1-GAL4AD, showed a similar transactivation capacity in both IPTG-induced and noninduced cells (Figure 2 and Supplementary Figure S2). A similar result was obtained when native ORE1 was tested (see later; Figure 3c).

**Functionality of Native Plant TFs in Yeast.** Above we characterized the transactivation capacity of plant TFs fused to the yeast GAL4 AD. However, considering that plant transcription activators typically harbor their own AD, we were interested to know whether or not native plant TFs can activate gene expression in the absence of a heterologous AD. To test this, we selected the NAC transcription factors JUB1 and ATAF1 which both showed strong transcription driver capacity in NLS-GAL4AD-TF and NLS-TF-GAL4AD fusions (Figure 2 and Supplementary Figure S2). As shown in Figure 3 panels a and b both native TFs activate *yEGFP* expression suggesting they are fully functional in yeast. However, transactivation capacity of the native TFs was somewhat weaker than that of the NLS-GAL4AD-TF and NLS-TF-GAL4AD fusions (Figure 2 and Supplementary Figure S2), indicating an additive effect of the plant and yeast (GAL4) ADs in supporting high transcriptional output (Figure 2 and Supplementary Figure S2).

The fact that no substantial *yEGFP* reporter activity was detected in the absence of the ORE1 transcription factor (Supplementary Figure S1), but relatively high reporter output was observed when the ORE1 CDS was coinserted into the yeast genome irrespective of the absence and presence of IPTG (which drives *ORE1* expression) (Figure 2 and Supplementary Figure S2), suggests that even leaky expression of *ORE1* is sufficient for considerable transcription activation output. This might be due to the GAL4 AD that is part of the chimeric ORE1 TFs, or caused by a motif intrinsic to ORE1. To distinguish between the two possibilities, we first tested the transactivation capacity of the native ORE1 protein. High *yEGFP* signal was detected in both, inducing and noninducing media (Figure 3c), indicating that ORE1 itself harbors a domain that causes the high background detected in non-inducing conditions.

We then generated ORE1-derived TFs by inserting the DBD of ORE1 (131 aa) downstream of the NLS and GAL4 AD, thereby creating the NLS-GAL4AD-DBD<sub>ORE1</sub> transcription factor (Figure 3d); we also created an NLS-DBD<sub>ORE1</sub>-GAL4AD regulator (Supplementary Figure S3). High *yEGFP* reporter activity was detected for both chimeric regulators in IPTG-induced cells, while *yEGFP* output was low in noninduced cells (Figure 3d and Supplementary Figure S3). Collectively, our results suggest that ORE1 harbors an activation domain that strongly affects transcriptional control in yeast.

**DNA-Binding Domains of Plant TFs Allow Tunable Transcriptional Output in Yeast.** We aimed at expanding the spectrum of transcriptional outputs of plant-derived ATFs by employing their DBDs fused to heterologous ADs. To this



**Figure 4.** Transcriptional output of ATFs harboring NAC transcription factor DBDs and GAL4 AD: (a) NLS-DBD<sub>JUB1</sub>-GAL4AD; (b) NLS-GAL4AD-DBD<sub>JUB1</sub>; (c) NLS-DBD<sub>ATAF1</sub>-GAL4AD; (d) NLS-GAL4AD-DBD<sub>ATAF1</sub>. Regulators were expressed from IPTG-inducible *GAL1* promoter. *yEGFP* output of the chromosomally integrated driver and effector constructs was determined in the absence and presence of inducer (IPTG). 1X, 2X, and 4X indicate one, two, or four copies of the TF binding site, respectively. Gray, noninduction medium; green, induction medium. N, negative control (*Pro*<sub>CYC1</sub> *min*-*yEGFP*); P, positive control (*Pro*<sub>TDH3</sub>-*yEGFP*). Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Asterisks indicate statistically significant difference from noninduction medium (Student's *t*-test; (\*)  $p < 0.05$ ; (\*\*)  $p < 0.01$ ). AU, arbitrary units. Full data are shown in [Supplementary Data S3](#).

end, we selected the NAC transcription factors JUB1 and ATAF1 which conferred robust activation of gene expression in yeast (Figure 2; [Supplementary Figure S2](#)). The two TFs share a conserved DBD of  $\sim 150$  amino acids, termed the NAC domain.<sup>31</sup> ATFs were created by inserting the DBD of JUB1 upstream (NLS-DBD<sub>JUB1</sub>-GAL4AD) or downstream (NLS-GAL4AD-DBD<sub>JUB1</sub>) of the yeast GAL4 AD. In the same way, chimeric ATAF1 TFs were produced (NLS-DBD<sub>ATAF1</sub>-GAL4AD and NLS-GAL4AD-DBD<sub>ATAF1</sub>). As before, the ATF's transactivation capacity was tested against one, two, or four copies of the binding sites inserted upstream of the *CYC1* minimal promoter that controls *yEGFP* reporter expression.

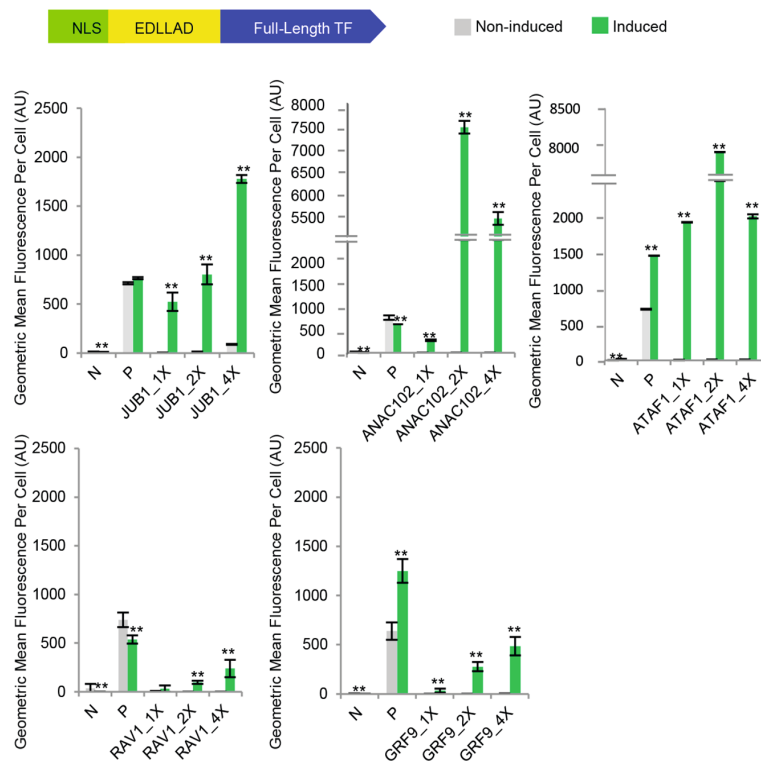
As seen in Figure 4, transcriptional output of all four NAC-derived regulators increased with the number of TF binding sites available in the target promoters, without a considerable effect on basal reporter gene expression. The activation level reached up to  $\sim 136$ -fold for the JUB1<sub>DBD</sub>, and up to  $\sim 68$ -fold for ATAF1<sub>DBD</sub>-derived ATFs (IPTG-induced and -non-induced) (Figure 4).

Collectively, our results show that transcriptional output of plant-derived ATFs can be tuned using different combinations of DNA-binding and activation domains, and copy numbers of DNA-binding sites. The data also demonstrate that ATFs

containing only the DBDs of plant TFs confer lower expression output than the entire TFs (Figure 2, Figure 3, and Figure 4). In general, removing either the plant AD (Figure 3d and Figure 4) or the yeast GAL4 AD (Figure 3a,b) resulted in ATFs with lowered transcription activation capacity.

**Plant EDLL Acts as a Strong Activation Domain in Yeast.** The *Arabidopsis* transcription factor ERF98/TDR1 harbors a strong, 24-amino-acid-long transactivation domain called EDLL that is conserved between different plant species.<sup>12</sup> However, to our knowledge, EDLL has not been reported as a functional AD in heterologous systems including yeast. We therefore generated additional ATFs by inserting the EDLL AD between the NLS and various TFs, namely JUB1, ANAC102, ATAF1, RAV1, and GRF9, which all showed moderate to high transcription activation property in the previous experiments (see above); expression of the chimeric NLS-EDLLAD-TF regulators was controlled by an IPTG-inducible promoter. Binding sites of the TFs were placed in one, two, or four copies upstream of the *CYC1* minimal promoter that controls *yEGFP* expression.

As shown in Figure 5, when combined with the NAC transcription factors JUB1, ANAC102, and ATAF1, EDLL acts as a strong activation domain in yeast, while fusing EDLL to



**Figure 5.** Transcriptional output of plant-derived ATFs fused to the EDLL AD. The plant transcription factors JUB1, ANAC102, ATAF1, RAV1, and GRF9 were fused N-terminally to the EDLL AD and their transcription activation capacity tested against one, two, or four copies of their DNA-binding sites (driver and reporter constructs chromosomally integrated). Expression of the synthetic regulators is under IPTG control; *yEGFP* reporter signal was determined in the absence and presence of inducer. 1X, 2X, and 4X indicate one, two, or four copies of the TF binding site, respectively. Gray, noninduction medium; green, induction medium. N, negative control (*Pro<sub>CYC1</sub> min-yEGFP*); P, positive control (*Pro<sub>TDH3</sub>-yEGFP*). Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Asterisks indicate statistically significant difference from noninduction medium (Student's *t*-test; (\*)  $p < 0.05$ ; (\*\*)  $p < 0.01$ ). AU, arbitrary units. Full data are shown in [Supplementary Data S4](#).

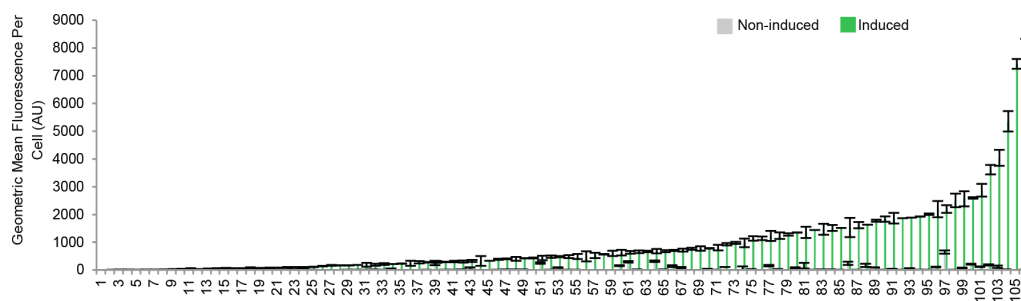
RAV1 or GRF9 less strongly enhanced transactivation capacity. Additionally, our results illustrate that basal activity (that is, ATF activity in the absence of IPTG) is influenced by the type of AD: most plant-derived EDLL ATFs tested here (JUB1, ANAC102, ATAF1, and RAV1) conferred significantly lower (or almost no) transcription output in noninduced cells than GAL4 AD containing ATFs.

Previously, the viral VP64 activation domain (VP64 AD) was embedded in synthetic TFs established on TALE and dCas9 frameworks.<sup>32,33</sup> Here, we generated plant-based ATFs carrying an NLS-VP64 AD at their C-terminus (TF-NLS-VP64AD) and tested their transactivation strength ([Supplementary Figure S4](#)). RAV1 or GRF9 fused to VP64 showed only marginal transactivation capacity ([Figure 5](#)). Transactivation capacity of JUB1 and ATAF1 fused to VP64 was not higher than those of the native TFs (JUB1, ATAF1; [Figure 2](#) and [3a,b](#)). In general, in yeast, fusion of VP64 AD to the C-terminus of plant TFs ([Supplementary Figure S4](#)) appeared to be less efficient in enhancing transcription output than fusions of VP64 AD to the C-terminus of TALEs (TALE\_NLS\_VP64 AD) or dCas9 (dCas9\_NLS\_VP64 AD).<sup>32,33</sup> For the ATFs generated here based on full-length plant TFs, activation levels of up to 307-, 2020-, and 105-fold (IPTG-induced versus noninduced) were observed upon fusion to the GAL4, EDLL, and VP64 activation domains, respectively ([Figure 2](#), [Figure 5](#), and [Supplementary Figure S4](#)).

Considering the strong transcription activation capacity conferred by EDLL AD when combined with NAC TFs in yeast, we tested its effect on transcriptional output when fused in one or two copies to JUB1. Fusion of EDLL AD to the C-terminus of JUB1 (NLS-JUB1-EDLL) created a chimeric TF that activates the *yEGFP* reporter to a level similar to that of the NLS-EDLL-JUB1 regulator ([Supplementary Figure S5a](#); compare with [Figure 5](#)). Importantly, however, fusing two copies of EDLL AD significantly increased the transcriptional output of the JUB1-derived ATF (NLS-JUB1-EDLL-EDLL) targeting four copies of its binding sites compared to the single EDLL AD fusion (NLS-JUB1-EDLL) ([Supplementary Figure S5b](#)).

Collectively, our data show that the plant EDLL motif acts as a strong transactivation domain in yeast ([Figure 5](#) and [Supplementary Figure S6](#)), suggesting it can be well implemented in synthetic biology ventures in this organism.

**A Wide Range of Transcriptional Outputs from Plant-Derived ATFs.** A central aim of our work was the establishment of a library of plant-derived TFs for future uses in synthetic biology projects dealing with yeast. In total, we established 106 different combinations of plants TFs, activation domains, and different numbers of the respective *cis*-regulatory motifs integrated upstream of the yeast *CYC1* minimal promoter. [Figure 6](#) illustrates the wide spectrum of transcriptional outputs we achieved by combining the different parts in



| Number | Plant-derived ATF/BS                     | Number | Plant-derived ATF/BS                | Number | Plant-derived ATF/BS                     |
|--------|--|--------|-------------------------------------|--------|--|
| 1      | NLS-ATAF1-GAL4AD/1X                      | 37     | NLS-EDLLAD-GRF9/2X                  | 71     | NLS-EDLLAD-JUB1/2X                       |
| 2      | NLS-DBD <sub>JUB1</sub> -GAL4AD/1X       | 38     | NLS-EDLLAD-ANAC102/1X               | 72     | NLS-JUB1-EDLLAD/4X                       |
| 3      | ATAF1/1X                                 | 39     | NLS-GAL4AD-ANT/1X                   | 73     | NLS-GRF7-GAL4AD/4X                       |
| 4      | NLS-GAL4AD-DBD <sub>JUB1</sub> /1X       | 40     | NLS-RAV1-GAL4AD/1X                  | 74     | NLS-DBD <sub>GRF7</sub> -GAL4AD/1X       |
| 5      | NLS-DBD <sub>ATAF1</sub> -GAL4AD/1X      | 41     | NLS-JUB1-GAL4AD/1X                  | 75     | NLS-GAL4AD-DBD <sub>JUB1</sub> /4X       |
| 6      | NLS-GAL4AD-DBD <sub>ATAF1</sub> /1X      | 42     | NLS-GAL4AD-RAV1/4X                  | 76     | NLS-GAL4AD-GRF7/4X                       |
| 7      | NLS-ANT-GAL4AD/1X                        | 43     | JUB1/4X                             | 77     | ATAF1/4X                                 |
| 8      | ANAC102-NLS-VP64AD/1X                    | 44     | NLS-GAL4AD-GRF9/1X                  | 78     | NLS-GAL4AD-ANAC102/2X                    |
| 9      | JUB1-NLS-VP64AD/1X                       | 45     | NLS-GAL4AD-GRF9/2X                  | 79     | NLS-GRF9-GAL4AD/2X                       |
| 10     | NLS-DBD <sub>ATAF1</sub> -GAL4AD/2X      | 46     | ANAC102-NLS-VP64AD/2X               | 80     | NLS-ANT-GAL4AD/2X                        |
| 11     | NLS-EDLLAD-RAV1/1X                       | 47     | NLS-GAL4AD-DBD <sub>JUB1</sub> /2X  | 81     | NLS-DBD <sub>GRF7</sub> -GAL4AD/2X       |
| 12     | JUB1-NLS-VP64AD/2X                       | 48     | NLS-DBD <sub>ATAF1</sub> -GAL4AD/4X | 82     | NLS-DBD <sub>ATAF1</sub> -linker-EDLL/4X |
| 13     | NLS-EDLLAD-GRF9/1X                       | 49     | NLS-RAV1-GAL4AD/4X                  | 83     | NLS-GAL4AD-DBD <sub>ATAF1</sub> /4X      |
| 14     | ATAF1-NLS-VP64AD/2X                      | 50     | NLS-GAL4AD-GRF9/4X                  | 84     | NLS-JUB1-EDLLAD/2X                       |
| 15     | RAV1-NLS-VP64AD/1X                       | 51     | JUB1-NLS-VP64AD/4X                  | 85     | NLS-JUB1-EDLLAD-EDLLAD/2X                |
| 16     | NLS-DBD <sub>JUB1</sub> -linker-EDLL/1X  | 52     | NLS-GAL4AD-ANAC102/1X               | 86     | NLS-DBD <sub>GRF7</sub> -GAL4AD/4X       |
| 17     | NLS-GAL4AD-RAV1/1X                       | 53     | NLS-GAL4AD-DBD <sub>GRF7</sub> /2X  | 87     | NLS-GAL4AD-ANAC102/4X                    |
| 18     | GRF9-NLS-VP64AD/2X                       | 54     | NLS-DBD <sub>JUB1</sub> -GAL4AD/2X  | 88     | ATAF1/2X                                 |
| 19     | NLS-GRF7-GAL4AD/1X                       | 55     | NLS-EDLLAD-GRF9/4X                  | 89     | NLS-EDLLAD-JUB1/4X                       |
| 20     | JUB1/1X                                  | 56     | NLS-GRF9-GAL4AD/1X                  | 93     | NLS-GAL4AD-ATAF1/1X                      |
| 21     | NLS-GAL4AD-DBD <sub>ATAF1</sub> /2X      | 57     | NLS-EDLLAD-JUB1/1X                  | 94     | NLS-GAL4AD-JUB1/2X                       |
| 22     | NLS-EDLLAD-RAV1/2X                       | 58     | NLS-GRF7-GAL4AD/2X                  | 92     | NLS-DBD <sub>ATAF1</sub> -linker-EDLL/2X |
| 23     | RAV1-NLS-VP64AD/2X                       | 59     | NLS-JUB1-EDLLAD/1X                  | 93     | NLS-JUB1-EDLLAD-EDLLAD/4X                |
| 24     | DBD <sub>JUB1</sub> -NLS-VP64/4X         | 60     | NLS-GAL4AD-DBD <sub>GRF7</sub> /4X  | 94     | NLS-EDLLAD-ATAF1/1X                      |
| 25     | GRF9-NLS-VP64AD/1X                       | 61     | NLS-GAL4AD-ANT/2X                   | 95     | NLS-EDLLAD-ATAF1/4X                      |
| 26     | NLS-GAL4AD-JUB1/1X                       | 62     | NLS-GAL4AD-RAV1/2X                  | 96     | NLS-GAL4AD-JUB1/4X                       |
| 27     | NLS-JUB1-EDLLAD-EDLLAD/1X                | 63     | ANAC102-NLS-VP64AD/4X               | 97     | NLS-ATAF1-GAL4AD/4X                      |
| 28     | GRF9-NLS-VP64AD/4X                       | 64     | NLS-GAL4AD-ANT/4X                   | 98     | ATAF1-NLS-VP64AD/4X                      |
| 29     | NLS-DBD <sub>JUB1</sub> -linker-EDLL/2X  | 65     | NLS-GAL4AD-GRF7/2X                  | 99     | NLS-GAL4AD-ATAF1/2X                      |
| 30     | NLS-DBD <sub>ATAF1</sub> -linker-EDLL/1X | 66     | NLS-JUB1-GAL4AD/4X                  | 100    | NLS-ANT-GAL4AD/4X                        |
| 31     | RAV1-NLS-VP64AD/4X                       | 67     | NLS-DBD <sub>JUB1</sub> -GAL4AD/4X  | 101    | NLS-GAL4AD-ATAF1/4X                      |
| 32     | NLS-GAL4AD-DBD <sub>GRF7</sub> /1X       | 68     | NLS-GRF9-GAL4AD/4X                  | 102    | NLS-ATAF1-GAL4AD/2X                      |
| 33     | JUB1/2X                                  | 67     | NLS-DBD <sub>JUB1</sub> -GAL4AD/4X  | 103    | NLS-JUB1-GAL4AD/2X                       |
| 34     | NLS-DBD <sub>JUB1</sub> -linker-EDLL/4X  | 68     | NLS-GRF9-GAL4AD/4X                  | 104    | NLS-EDLLAD-ANAC102/4X                    |
| 35     | NLS-GAL4AD-GRF7/1X                       | 69     | ATAF1-NLS-VP64AD/1X                 | 105    | NLS-EDLLAD-ANAC102/2X                    |
| 36     | NLS-EDLLAD-RAV1/4X                       | 70     | NLS-RAV1-GAL4AD/2X                  | 106    | NLS-EDLLAD-ATAF1/2X                      |

**Figure 6.** Library of genome-integrated, plant-based ATFs. The CDSs of plant TFs or their DBDs were fused in different combinations with GAL4, VP64, or EDLL activation domains and their transactivation capacity was tested against the TF's respective binding sites inserted in one (1X), (2X), and (4X) copies upstream of the *CYC1* minimal promoter driving *yEGFP* reporter expression. Expression of ATFs was controlled by the IPTG-inducible *GALI1* promoter and *yEGFP* output was tested in the absence and presence of IPTG (20 mM). Each number (1–106) indicates one type of ATF, as listed in the table. Gray, noninduction medium; green, induction medium. Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. AU, arbitrary units. Full data are shown in [Supplementary Data S5](#).

various ways and integrating them into the yeast genome. In addition to the combinations reported above, the library includes further parts modifications, namely the addition of a 40-aa linker in front of the EDLL AD in NLS-DBD<sub>TF</sub>-linker-EDLL regulators (for details see legend to [Figure 6](#)) which added to the range of available transcriptional outputs provided by the plant-based ATFs.

Of note, several of the ATFs elicited considerably higher *yEGFP* reporter output than the control that expresses *yEGFP* from the strong constitutive yeast *TDH3* promoter (*Pro<sub>TDH3</sub>-yEGFP*). The strongest transcription activation capacity we observed for NAC transcription factor-derived ATFs, in particular NLS-JUB1-GAL4AD/2X, NLS-EDLL-ATAF1/2X, NLS-EDLL-ANAC102/2X, and NLS-EDLL-ANAC102/4X,

which gave  $\sim$ 6- and  $\sim$ 10-fold stronger reporter output than the positive control.

**Transcriptional Output of RAV1-Based ATF Implemented in Centromeric Plasmids.** After successful generation of a library of plant-derived ATFs, we next investigated the transcriptional output of an ATF encoded from plasmids. We selected the NLS-RAV1-GAL4AD transcription factor, which showed a midrange fluorescence output of target gene expression when chromosomally integrated ([Supplementary Figure S2](#)), and employed the same regulatory system implemented in centromeric plasmid pGN006 containing a *CEN/ARS* low-copy replication origin. Upon induction, the ATF targets two copies of the RAV1 binding site upstream of the *CYC1* minimal promoter in the same vector ([Supplementary Figure S7a](#)). Despite a 2–5 times higher copy number



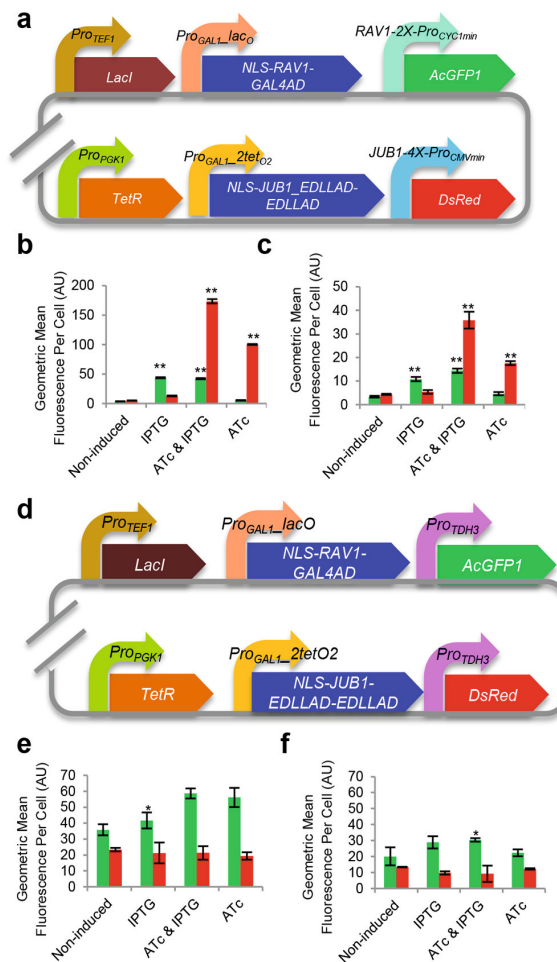
of *CEN/ARS* plasmids compared to chromosomally integrated cassettes,<sup>34</sup> we observed a ~40% lower yEGFP reporter output in the centromeric plasmid than in the genome-integrated framework (Supplementary Figure S7b).

In a further experiment, we placed NLS-RAV1-GAL4AD downstream of the ATc-inducible *GAL1* promoter also present in pGN006 (Supplementary Figure S7a) and tested activation of the same yEGFP reporter as above after induction with ATc. As the *TEF1* promoter driving *Lacl* expression and the *PGK1* promoter driving *TetR* expression in pGN006 have similar strengths,<sup>35</sup> and as the concentrations of IPTG (20 mM) or ATc (1.08  $\mu$ M) used in our experiments lead to full induction of IPTG- and ATc-inducible *GAL1* promoters, respectively,<sup>5,13</sup> the yEGFP fluorescence output of the NLS-RAV1-GAL4AD transcription factor was expected to be similar. Indeed, fluorescence outputs were comparable for both strains (Supplementary Figure S7b). The results demonstrate that plant TFs can be employed as orthogonal drivers of gene expression from their cognate binding sites when implemented in yeast centromeric plasmids. Additionally, we observed that ~50% of the cells carrying centromeric plasmids produced yEGFP, while in the case of genome-integrated constructs ~80% of the cells showed yEGFP signal (Supplementary Figure S7c).

#### Expression of Two Plant-Derived ATFs from Plasmids.

We have recently reported the establishment of a highly efficient and versatile multigene assembly technology, called AssemblX, which allows the rapid generation of complex DNA constructs for use in synthetic biology.<sup>36</sup> However, for the full exploitation of the AssemblX and other genetic circuit construction tools,<sup>36–38</sup> well-defined (orthogonal) transcription regulatory systems (networks and their components) are needed. We therefore tested whether two different plant-derived ATFs, encoded on the same plasmid, can be employed to drive gene expression from their cognate binding sites in yeast. To minimize potential interference of the ATFs with each other's binding sites, we chose regulators with divergent *cis*-regulatory sequences, namely RAV1 and JUB1. We inserted the coding sequences of the chimeric transcription factors NLS-RAV1-GAL4AD and NLS-JUB1-EDLL-EDLL into plasmids pGN008 and pGN009, harboring *CEN/ARS* low-copy and  $2\mu$  high-copy replication origins, respectively. In both vectors, expression of NLS-RAV1-GAL4AD and NLS-JUB1-EDLL-EDLL is independently controlled by IPTG- or ATc-inducible *GAL1* promoters, respectively (Figure 7a). NLS-RAV1-GAL4AD-mediated transcriptional control was tested against two RAV1 binding sites inserted upstream of the *CYC1* minimal promoter, with AcGFP1 (*Aequorea coerulescens* green fluorescent protein) as reporter. Control executed by NLS-JUB1-EDLL-EDLL was tested against four JUB1 binding sites placed upstream of the *CMV* minimal promoter, with DsRed (*Discosoma* sp. red fluorescent protein) as reporter. The combinations of ATFs and binding sites were chosen to represent regulatory systems with moderate (RAV1; Supplementary Figure S2) and high (JUB1; Supplementary Figure S5b) transcriptional output.

Simultaneous induction of NLS-RAV1-GAL4AD and NLS-JUB1-EDLL-EDLL (by IPTG and ATc) results in an increase of both reporter (AcGFP1 and DsRed) intensities. Overall, we observed higher fluorescence output for cells harboring pGN009 (Figure 7b) than pGN008 (Figure 7c), likely due to higher copy number of the pGN009 plasmid ( $2\mu$  origin), consistent with previous studies.<sup>39,40</sup>



**Figure 7.** Coexpression of plant-derived ATFs integrated in both episomal and centromeric plasmids. (a) Schematic representation of plasmids allowing IPTG- and ATc-inducible expression of NLS-RAV1-GAL4AD and NLS-JUB1-EDLL-EDLL, respectively. Expression of the reporter gene *AcGFP1* is controlled by two RAV1 binding sites (2X) inserted upstream of the *CYC1* minimal promoter, while expression of the reporter gene *DsRed* is controlled by four JUB1 binding sites (4X) inserted upstream of the *CMV* minimal promoter. Regulator and reporter constructs were integrated in both, centromeric (pGN008, *CEN/ARS* origin) and episomal (pGN009,  $2\mu$  origin) plasmids. (b) Green (AcGFP1) and red (DsRed) fluorescence outputs measured by flow cytometry in the absence and presence of the appropriate inducer(s) in pGN009 vector backbones. (c) Green (AcGFP1) and red (DsRed) fluorescence outputs measured in the absence and presence of the appropriate inducer(s) in pGN008 vector backbones. (d) Scheme of positive control plasmids, which allow constitutive expression of both reporter genes (*Pro<sub>TDH3</sub>*-AcGFP1 and *Pro<sub>TDH3</sub>*-DsRed) independent of the expression level of the ATFs. The constructs were established in both, pGN008 (*CEN/ARS*) and pGN009 ( $2\mu$ ) vector backbones. (e) Green (AcGFP1) and red (DsRed) fluorescence outputs measured in pGN009 vector backbones. (f) Green (AcGFP1) and red (DsRed) fluorescence outputs without or with the appropriate inducer(s) in pGN008 vector backbones. Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Asterisks indicate statistically significant differences from the respective controls (i.e., IPTG-containing vs IPTG-free medium, and ATc-containing vs ATc-free

Figure 7. continued

medium; Student's *t*-test; (\*)  $p < 0.05$ ; (\*\*)  $p < 0.01$ . AU, arbitrary units. Full data are shown in [Supplementary Data S6](#).

As a control we also tested fluorescent reporter output from vectors expressing NLS-RAV1-GAL4AD and NLS-JUB1-EDLL-EDLL from IPTG- and ATc-inducible promoters (as above), but expressing the AcGFP1 and DsRed reporters from the strong and constitutive yeast *TDH3* promoter (Figure 7d)<sup>41</sup> not containing the plant *cis*-regulatory motifs (*Pro<sub>TDH3</sub>-AcGFP1* and *Pro<sub>TDH3</sub>-DsRed*); the respective vectors are pGN009-PosCon ( $2\mu$ ; based on pGN009) (Figure 7e) and pGN008-PosCon (*CEN/ARS*; based on pGN008) (Figure 7f). Notably, AcGFP1 reporter output was similar in yeast cells expressing *AcGFP1* from the *TDH3* promoter and cells expressing *AcGFP1* from the *CYC1* minimal promoter harboring the RAV1 binding sites (after IPTG induction) (Figure 7b,c,e,f). In contrast, DsRed reporter output in cells expressing *DsRed* from the *CMV* minimal promoter harboring the JUB1 binding sites was considerably higher (after ATc induction) than in cells expressing the reporter from the *TDH3* promoter (Figure 7b,c,e,f).

Finally, we tested NLS-RAV1-GAL4AD- and NLS-JUB1-EDLL-EDLL-driven AcGFP1 and DsRed reporter outputs from plasmids harboring the *CYC1* and *CMV* promoters, respectively, but lacking the RAV1- and JUB1 *cis*-regulatory motifs (negative controls, *Pro<sub>CYC1 min</sub>-AcGFP1* and *Pro<sub>CMV min</sub>-DsRed*). The respective vectors are pGN008-NegCon, with a *CEN/ARS* origin, and pGN009-NegCon with a  $2\mu$  origin (Supplementary Table S4). No considerable fluorescence output was observed in the absence or presence of the appropriate inducer(s) (data not shown).

Collectively, as in the chromosomally inserted versions, JUB1 appears as an excellent tool for driving strong gene expression from synthetic promoters (harboring JUB1 binding sites), similar to other NAC transcription factors, namely ANAC102 and ATAF1. Our data therefore show that plant TFs, in particular NACs, can be faithfully employed as individually acting orthogonal drivers in yeast.

**Growth Effects.** To test the effect of high transcription factor expression on cell growth, we selected six JUB1, ATAF1-, and ANAC102-derived ATFs whose transcriptional output, determined as yEGFP fluorescence, was 6- to 10-fold stronger after IPTG induction than that of the yeast *TDH3* promoter. As expected, in the presence of ATF-encoding plasmids, growth was generally reduced compared to cells not containing an ATF. However, in the majority of the cases (five out of six ATFs tested), reduction of growth was similar to cells expressing the yEGFP reporter from the strong *TDH3* promoter (Supplementary Figure S8). Thus, control of gene expression by plant-derived ATFs to achieve high transcription output is excellently suited for synthetic biology applications in yeast.

## DISCUSSION

Artificial transcription factors (ATFs) are important tools for establishing orthogonal transcriptional regulatory systems in heterologous hosts and, thus, are central to synthetic biology applications. The yeast *Saccharomyces cerevisiae* is an important model organism for molecular and cell biological research, while it also plays an important role in a variety of industrial

applications.<sup>2</sup> Furthermore, *S. cerevisiae* is increasingly being recognized as an excellent model organism for synthetic biology research.<sup>2,14</sup>

Two important aspects of cellular engineering of yeast (or other organisms) include: (i) the possibility to assemble multicomponent DNA constructs harboring genes and their parts, and (ii) the availability of control elements, such as TFs and their *cis*-regulatory elements that can be faithfully combined for establishing the core of orthogonal regulatory systems. Thanks to recent progress researchers have now access to various gene assembly technologies that allow them to build even complex genetic constructs in relatively short time,<sup>36,42</sup> although considerable limitations still often exist.<sup>3,5,14</sup> On the other hand, components for orthogonal regulation of cellular activities at diverse levels (DNA replication and cell proliferation, transcription, translation, metabolic channeling, and others) are still not sufficiently well established for a large number of cellular processes in different organisms.

TFs from plants and animals, including humans, are often expressed in yeast in the frame of one-hybrid and two-hybrid screens, to identify TFs binding to defined promoters (or *cis*-regulatory motifs embedded in them) or proteins interacting with the TFs; such experiments had been very successful in the past, demonstrating that alien (including plant) TFs are faithfully folded and generally functional in yeast. However, plant TFs have so far not much been employed for setting up orthogonal regulatory systems within the realm of synthetic biology in yeast. This is perhaps surprising considering that yeast has only about 300 TFs,<sup>43</sup> while higher plants have 2000 TFs or more,<sup>44</sup> providing ample opportunities for constructing orthogonal synthetic regulatory transcription networks in yeast. Several of the plant TF families, including the NAC TFs we tested here for the control of gene expression in yeast, are absent from nonplant pro- and eukaryotes. Such TFs might, therefore, be particularly well suited for establishing orthologous gene regulatory cascades and networks in yeast (or other heterologous) systems.

Here, we selected TFs from the higher plant *Arabidopsis thaliana* with previously established DNA-binding motifs, including NAC TFs (Supplementary Table S1), for functional implementation in yeast. We tested their transcription activation potential toward their binding sites, either in native format or in fusion with different transactivation domains from yeast itself (GAL4 AD) or from heterologous hosts (VP64 from herpes simplex virus, or EDLL from *Arabidopsis*). In total, we tested 106 different combinations of ATFs and binding sites after chromosomal integration and observed a wide spectrum of inducible transcriptional outputs, ranging from as low as 1 to more than 2000-fold (Figure 6).

Major findings we made are as follows: (i) Plant TFs are excellent tools for establishing inducible transcriptional units in yeast; they can be used for regulating gene expression over a wide range of magnitudes. (ii) Members of the plant NAC family are particularly well suited for establishing orthogonal transcription units in yeast. This is in accordance with the fact that NAC TFs are entirely absent from organisms outside the plant kingdom.<sup>45</sup> (iii) The plant EDLL activation domain, chosen from the *Arabidopsis* ERF98 TF, is well suited for enhancing transcription activation output when fused to heterologous TFs and expressed in yeast. Particularly strong transcription output was observed when EDLL was implemented within the frame of ATFs derived from the NAC TFs JUB1, ANAC102, and ATAF1 (Figure 5). Important also, in

several cases (e.g., for ATFs based on ANAC102, ATAF1, and RAV1) we observed a reduction in background reporter activity in the absence of IPTG in fusions containing EDLL, suggesting a tighter control of transcription activation. (iv) Transcription output achieved by plant-derived ATFs can be considerably, i.e. 6- to 10-fold, higher than that of the strong and constitutive yeast *TDH3* promoter. This is an important finding as it may allow boosting the expression of genes of interest to unprecedented high levels in this organism, which may, for example, support the controlled up-regulation of secondary metabolite biosynthesis. (v) Plant transcription factors and their cognate binding sites can be employed for building independently acting regulators, potentially allowing to set up complex control networks by combining different TF-binding site combinations within the frame of one plasmid or a synthetic chromosome. As a proof-of-concept we implemented two different plant-derived ATFs, namely NLS-RAV1-GAL4AD and NLS-JUB1-EDLL-EDLL, into centromeric and episomal plasmids and demonstrated independent transcriptional control toward their respective synthetic promoters harboring RAV1 and JUB1 binding sites, respectively.

The majority of our chosen plant TFs with considerable reporter output (namely JUB1, ORE1, ATAF1, ANAC102, GRF7, and GRF9) carry their native AD at the C-terminus. For ATFs derived from those TFs obtained by adding a nonplant AD (i.e., the GAL4 AD) to their N-terminus, leading to AD-TF-AD fusions, we observed that transcriptional output generally increased when the number of binding sites increased (Figure 2). This correlation of reporter output with the number of binding sites was less evident for TF-AD-AD fusions (which carry the GAL4 AD at the C-terminus of the ATFs; Supplementary Figure S2). Collectively, our observations therefore indicate that keeping a larger distance between two different ADs (the native plant AD and the heterologous yeast AD) in the primary structure of the ATFs generally favors transcriptional output with increasing number of binding sites, at least in the plant TFs tested here. This may be due to higher flexibility of the arrangement of the plant and yeast ADs around the DNA binding domain, thereby giving a higher probability to interact with the general transcription activation machinery to favor transcriptional activity from the promoter.

In several cases, we observed the highest transcriptional output for TF-AD-AD fusions with two DNA binding sites (Supplementary Figure S2), which tended to exceed that of AD-TF-AD fusions with two or four binding sites (Figure 2). Thus, for the planning of other ATFs in the future, the experimenter might choose between two options: (i) If highest transcription output is wanted, option 1 (TF-AD-AD with two binding sites) might be the preferred choice, in particular when ATFs are based on plant NAC TFs. (ii) If a gradual increase in transcription output is wanted, option 2 (AD-TF-AD with one, two, and four binding sites) appears to be the better option in most cases.

Of general interest in synthetic biology projects involving TFs is their specificity with respect to the genes directly controlled by them. We therefore searched the yeast genome for potential binding sites of the plant TFs employed in our study; the results are presented in Supplementary Table S1. With respect to TFs that show strong transcription activation capacity in yeast (i.e., NAC TFs), we found that binding sites for the NAC TFs are either absent (0 hits) or rare in the yeast genome. This is likely due to their relatively long (bipartite) binding site and the fact that NAC TFs are naturally absent

from yeast (NAC TFs are plant-specific). Similarly, binding sites for ANT and RAV1, which led to moderately high reporter output, are relatively long and absent (0 hits) from the yeast genome (Supplementary Table S1). The plant TFs GRF7 and GRF9 led to moderate reporter output in yeast; their binding sites are relatively short and more frequent in the yeast genome, which is not unexpected. Finally, binding sites for WRKY6 (6 bp), DOF1 (5 bp), and MYB61 (6 bp) are abundant in the yeast genome. However, these three plant TFs did not activate the synthetic promoters containing up to four copies of the binding sites (Figure 2 and Supplementary Figure S2), strongly indicating they cannot properly interact with the basic transcription machinery to activate transcription in yeast.

Taking into account that plant NAC TFs have a strong transactivation capacity in yeast, that their binding sites are virtually absent from the yeast genome, and that numerous NAC TFs are encoded by plant genomes we strongly suggest them as excellent candidates for the establishment of orthologous transcription regulatory circuits in synthetic biology projects in the future.

Transcription factors often interact with other TFs of either the same or a different family, or with proteins that do not directly bind DNA (transcription regulators), to form heterodimers or higher-order regulatory complexes.<sup>46</sup> The plant NAC TFs are well-known for their capacity to form heterodimers and to interact with non-TF proteins through their highly divergent C-terminal segments.<sup>47</sup> NACs represent one of the largest TF family in higher plants, with typically 100–150 members in each species,<sup>47</sup> providing a large repertoire for a further expansion of synthetic gene expression regulators for implementation in yeast. In addition, our increasing knowledge about non-TF interactors of NACs<sup>5</sup> provides a currently unexplored resource for potential integration into regulatory protein complexes or for interfacing synthetic expression networks with metabolic or other cellular elements.

Although we used yEGFP as the reporter for transcription output in our study, we envisage that a wide range of transcriptional activities will also be achieved with other open reading frames inserted downstream of the synthetic promoters, thereby facilitating the construction of more complex regulatory systems. In addition, although we used IPTG and ATc to drive ATF expression in our studies, other chemical inducers or even light-controlled molecular switches may be implemented instead.<sup>48</sup> Furthermore, the synthetic, plant-based ATFs and transcription control modules we report here may be integrated into existing cloning frameworks for synthetic biology, for example, the AssemblX toolbox.<sup>36</sup>

Another aspect for further refinement is the following: so far, we tested the functionality of the plant-derived *cis*-regulatory elements only within the frame of the yeast *CYC1* minimal promoter (with the exception of the two-TF plasmid, where we also used the *CMV* minimal promoter). However, a potentially large set of other minimal promoters from either yeast or plant origin can be established and modified by altering the number of ATF binding sites within such regulatory parts. Given the wide range of transcription output intensities observed here for TF binding sites fused to the *CYC1* minimal promoter, we expect that a similar broad range of expression outputs can be achieved by integrating the ATF's binding sites into other (minimal) promoters. This will allow establishing a vast spectrum of controllable transcription regulatory systems,



with low to very high expression outputs, strongly supporting future applications in synthetic biology.

Previously, we reported the construction of a library of 40 constitutive yeast promoters, each followed by a cloning site and the cognate terminator, for controlled expression of heterologous coding sequences in yeast.<sup>36</sup> Although this library of regulatory sequences allows gene expression over a wide range of activities, control over them is executed by endogenous yeast TFs that typically maneuver a suite of different genes within their respective gene regulatory networks. Thus, the control of heterologous genes occurs within the frame of the regulatory network governed by the yeast TF, which may not be wanted in many cases. The plant-derived orthogonal ATFs presented here, together with their cognate binding sites, strongly increase the repertoire of transcriptional regulatory modules for use in synthetic biology in yeast. As for establishing the library of the 106 regulatory modules, we only included plant binding sites that were largely inactive in the absence of the corresponding TF, we expect minimal interference from the yeast transcription control machinery. Our ATF/*cis*-regulatory library therefore represents an excellent starting point for further improvements of orthogonal regulatory systems in yeast.

## CONCLUSIONS

We generated a set of 106 synthetic transcription regulatory units based on plant TFs and their cognate binding sites, covering a wide range of transcriptional outputs, with up to 2000-fold induction level. Some transcriptional control units provide expression outputs which are 6- to 10-fold stronger than that of the widely used strong yeast *TDH3* promoter. The diverse regulatory units were assembled from DBDs or full-length coding sequences of plant TFs, fusions to different ADs (from yeast, virus, and plant origin), and one, two, and four copies of their binding sites. We furthermore showed that plant-derived ATFs can be implemented as independently acting control units in extrachromosomal plasmids with high- and low-copy replication origins facilitating the future generation of synthetic chromosomes. We expect that the transcription units established here can also be employed in other organisms such as mammalian cells.

## MATERIALS AND METHODS

**General.** Plant TFs included in this work as well as their core binding sites are given in [Supplementary Table S1](#). The core binding sites (shown in bold in the table) are embedded in a longer, 19–40 bp long DNA sequence. The extended DNA fragments used to make the constructs were taken from EMSA or binding site selection assays previously employed to demonstrate binding of the respective TF to the DNA, or transactivation assays. The chosen DNA fragments typically gave the highest binding affinity/transactivation capacity in the published papers.

Plasmids were constructed by Gibson assembly<sup>49</sup> and SLiCE cloning.<sup>50,51</sup> All constructs were confirmed by sequencing (Eurofins Genomics, Ebersberg, Germany, or LGC Genomics, Berlin, Germany).

**Bacterial and Yeast Strains.** Plasmids were transformed into competent *Escherichia coli* DH5 $\alpha$  cells or into NEB5 $\alpha$  or NEB10 $\beta$  cells (New England Biolabs, Frankfurt am Main, Germany). *Saccharomyces cerevisiae* strain YPH500 (*MATA*, *ura3-52*, *lys2-801*, *ade2-101*, *trp1 $\Delta$ D63*, *his3 $\Delta$ 200*, *leu2 $\Delta$ 1*)

(ATCC: #76626) was used as host in all yeast experiments. Preparation of competent yeast cells and genetic transformation of plasmids or linearized DNA fragments were done using the LiAc/SS carrier DNA/PEG method.<sup>52</sup> Culturing yeast cells and confirmation of transformation events were done using *HIS3*, *TRP1*, or *URA3* selectable markers and via colony PCR followed by sequencing.

**Construction of Synthetic Promoters.** To construct synthetic promoters, we annealed single-stranded oligonucleotides containing one, two, or four binding sites of plant TFs and inserted them into *XbaI/SalI*-digested reporter plasmid, upstream of the *CYC1* minimal promoter. Reporter plasmids ([Supplementary Methods](#)) were linearized with *PmeI* (present within the *URA3* homology buffer) and transformed into yeast strain YPH500. Integration takes place at the *ura3-52* locus of the yeast genome. Positive clones were selected on SD-His medium. Primer sequences and reporter plasmids are given in [Supplementary Table S2](#).

### Construction of Plant-Derived ATF Expression Clones.

Coding sequences of plant TFs, or their DBDs, were amplified by PCR from *Arabidopsis thaliana* Col-0 leaf cDNA (for primer sequences see [Supplementary Table S3](#)). Expression plasmids ([Supplementary Methods](#)) were digested with *BamHI* and *AgeI* and assembled with PCR fragments resulting in plant-derived ATFs harboring a C- or N-terminal transactivation domain (AD); except for pGN005D, which does not encode an AD. The resulting expression plasmids were linearized with *AatII* and transformed into yeast strain YPH500 containing the reporter cassette of the corresponding target promoter. Integration of expression plasmids (containing plant-derived ATFs) takes place in the already existing reporter cassette in the yeast genome. Positive clones were selected on SD-Trp/-His medium. Subsequently expression of the chimeric TFs was controlled by an IPTG-inducible promoter, and their transactivation capacity was tested against one, two, or four TF binding sites placed upstream of the *CYC1* minimal promoter that controls *yEGFP* expression.

### Construction of Yeast Plasmids for Expression of Two Plant-Derived ATFs.

Synthetic promoters harboring two RAV1 binding sites (primers GN055/GN056; on pGN005B-2xBS-RAV1; [Supplementary Table S2](#)) were cloned into *AvrII/SalI*-digested pGN006, pGN008, and pGN009 ([Supplementary Methods](#)). Synthetic promoters harboring four JUB1 binding sites (primers GN057/GN058; on pGN005B-4xBS-JUB1; [Supplementary Table S2](#)) were cloned into *PfI/PI/PmeI*-digested pGN008 and pGN009. The CDS of the NLS-RAV1-GAL4AD transcription factor (primers GN059/GN060; on pFM003B-RAV1; [Supplementary Table S3](#)) was inserted into pGN006, between *AgeI* and *RsrII* sites downstream of the ATc-inducible promoter. In addition, the NLS-RAV1-GAL4AD transcription factor (primers GN061/GN062; on pFM003B-RAV1) was inserted between *SexAI* and *RsrII* sites, downstream of the IPTG-inducible promoter in plasmids pGN006, pGN008, and pGN009. The CDS of NLS-JUB1-EDLL-EDLL (primers GN063/GN064; on pGN003C-JUB1; [Supplementary Table S3](#)) was inserted between *AgeI* and *RsrII* sites, downstream of the ATc-inducible promoter in plasmids pGN008 and pGN009.

The positive control vectors pGN008-PosCon and pGN009-PosCon ([Supplementary Table S4](#)) were constructed as follows: pGN008 and pGN009 were digested with *SalI* and *BamHI* to remove the *CYC1* minimal promoter, and PCR-amplified *TDH3* promoter (primers GN065/GN066; on yeast

*Saccharomyces cerevisiae* BY4741 DNA) was inserted (*Pro<sub>TDH3</sub>-AcGFP1*). Subsequently, the resulting plasmids were digested with *PmeI* and *BsaBI* to remove the *CMV* minimal promoter, and PCR-amplified *TDH3* promoter (primers GN067/GN068; on yeast BY4741 DNA) was inserted (*Pro<sub>TDH3</sub>-DsRed*). As negative controls, pGN008 and pGN009 harboring minimal promoters without plant TF binding sites were used (*Pro<sub>CYC1 min</sub>-AcGFP1* and *Pro<sub>CMVmin</sub>-DsRed*). The primers are listed in [Supplementary Table S4](#). Centromeric or episomal plasmids were directly transformed into yeast strain YPH500. Positive clones were selected on SD-Trp medium.

#### Induction Experiments, Flow Cytometry and Data

**Analysis.** To determine the yEGFP fluorescence output in the absence of plant TFs, single colonies of yeast reporter strains were inoculated into 500  $\mu$ L SD-His medium in 48-well deep-well plates. Plates were incubated for 18–24 h at 30 °C in a rotary shaker at 230 rpm. The precultures were used to inoculate main cultures in 500  $\mu$ L of YPDA (containing 2% glucose) to an  $OD_{600} \sim 0.1$ . Each reporter strain was inoculated in three technical replicates per experiment. Cells were grown at 30 °C for 14–16 h in a rotary shaker at 230 rpm. Samples were treated with cycloheximide at a final concentration of 500  $\mu$ g/mL to inhibit protein synthesis and analyzed using a BD FACSCalibur flow cytometer (BD Biosciences). yEGFP fluorescence values were obtained from a minimum of 10 000 cells in each sample. The geometric mean of the yEGFP fluorescence per cell was calculated using Flowing Software version 2.5.1 (<http://www.uskonaskel.fi/flowingsoftware>).

To determine the effect of plant-derived ATF on yEGFP fluorescence output, the same procedure as above was employed except that yeast precultures were grown in 500  $\mu$ L of SD -Trp/-His medium (synthetic drop-out media containing 2% glucose) and main cultures were grown in YPDA containing 2% glucose (noninducing medium) or YPDA containing 2% galactose and 20 mM IPTG (inducing medium). For AcGFP1 or/and DsRed fluorescence measurements, the same procedure as above was employed except that yeast precultures were grown in 500  $\mu$ L of SD-Trp medium containing 2% glucose and main cultures were grown in SD-Trp medium containing 2% glucose (noninducing medium) or 2% galactose and 20 mM IPTG with or without 1.08  $\mu$ M ATC (inducing media).

**Search for Transcription Factor Binding Sites in the Yeast Genome.** The presence of potential binding sites of plant TFs in the yeast genome was explored using the Pattern Matching tool available at the Saccharomyces Genome Database ([www.yeastgenome.org](http://www.yeastgenome.org)), using the yeast reference strain S288C and the setting “genoSc” (Complete *S. cerevisiae* Genome DNA). The results are shown in [Supplementary Table S1](#).

**Growth Assays.** Growth assays were done similarly to induction experiments, except that experimental cultures were inoculated to an  $OD_{600}$  of  $\sim 0.05$  and grown at 30 °C and 230 rpm in a rotary shaker.  $OD_{600}$  was measured after 8, 16, 24, 36, 48, and 64 h.

#### ■ ASSOCIATED CONTENT

##### 📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acssynbio.7b00094](https://doi.org/10.1021/acssynbio.7b00094).

Supplementary Figure S1. Binding site specificity of plant TFs in yeast. Supplementary Figure S2. Transcriptional

output of plant-derived NLS-TF-GAL4AD regulators. Supplementary Figure S3. Transcriptional output of NLS-DBD<sub>ORE1</sub>-GAL4AD. Supplementary Figure S4. Transcriptional output of plant-derived ATFs harboring VP64 AD. Supplementary Figure S5. Effect of EDLL AD on transcriptional output of JUB1. Supplementary Figure S6. Fold activation for plant-derived ATFs harboring GAL4 AD, EDLL AD or VP64 AD. Supplementary Figure S7. Transcriptional output of RAV1-derived ATF integrated in yeast plasmids. Supplementary Figure S8. Growth assays. Supplementary Figure S9. Plasmid maps (PDF)

Supplementary Table S1. Transcription factors from *Arabidopsis thaliana* employed for generating ATFs. Supplementary Table S2. Sequences of reporter plasmids and primers. Supplementary Table S3. Sequences of expression plasmids and primers. Supplementary Table S4. Sequences of centromeric and episomal plasmids and primers (XLSX)

Full data for figures and supplementary figures (XLSX)

#### ■ AUTHOR INFORMATION

##### Corresponding Author

\*E-mail: [bmr@uni-potsdam.de](mailto:bmr@uni-potsdam.de).

##### Author Contributions

G.N. and B.M.R. designed the experiments, with contributions from K.M. and F.M. S.B. provided information about the binding sites of transcription factors and advice for the analysis of transcription factor–DNA interactions. I.K. supported the ANAC102-related experiments. F.M. developed the pFM series of plasmids. G.N. performed the experiments and analyzed the data. G.N., S.B., and B.M.R. wrote the manuscript, which was proofread by all authors. All authors take full responsibility for the content of the paper.

##### Notes

The authors declare no competing financial interest.

#### ■ ACKNOWLEDGMENTS

We are thankful to Dr. Pamela Holzlohner and Dr. Mohammad A. Omidbakhshfard (University of Potsdam) for help with the flow cytometry measurements and for providing GRF9 binding site information, respectively. This research was funded by the Federal Ministry of Education and Research of Germany (BMBF; FKZ 031A172) and through a Ph.D. scholarship funding from the Potsdam Graduate School (PoGS, University of Potsdam).

#### ■ ABBREVIATIONS

AD, activation domain; ATF, artificial transcription factor; BS, binding site; CDS, coding DNA sequence; DBD, DNA-binding domain; NLS, nuclear localization signal; yEGFP, yeast enhanced green fluorescent protein

#### ■ REFERENCES

- (1) Borodina, I., and Nielsen, J. (2014) Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnol. J.* 9, 609–620.
- (2) Nielsen, J., and Keasling, J. D. (2016) Engineering Cellular Metabolism. *Cell* 164, 1185–1197.
- (3) Holtz, W. J., and Keasling, J. D. (2010) Engineering static and dynamic control of synthetic pathways. *Cell* 140, 19–23.

- (4) Blount, B. A., Weenink, T., and Ellis, T. (2012) Construction of synthetic regulatory networks in yeast. *FEBS Lett.* 586, 2112–2121.
- (5) Khalil, A. S., Lu, T. K., Bashor, C. J., Ramirez, C. L., Pyenson, N. C., Joung, J. K., and Collins, J. J. (2012) A synthetic biology framework for programming eukaryotic transcription functions. *Cell* 150, 647–658.
- (6) Kabadi, A. M., and Gersbach, C. A. (2014) Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression. *Methods* 69, 188–197.
- (7) Purcell, O., Peccoud, J., and Lu, T. K. (2014) Rule-based design of synthetic transcription factors in eukaryotes. *ACS Synth. Biol.* 3, 737–744.
- (8) Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., and Yokoyama, S. (2013) DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends Plant Sci.* 18, 267–276.
- (9) Sun, N., and Zhao, H. (2014) A single-chain TALEN architecture for genome engineering. *Mol. BioSyst.* 10, 446–453.
- (10) Redden, H., and Alper, H. S. (2015) The development and characterization of synthetic minimal yeast promoters. *Nat. Commun.* 6, 7810.
- (11) Feng, J., Jester, B. W., Tinberg, C. E., Mandell, D. J., Antunes, M. S., Chari, R., Morey, K. J., Rios, X., Medford, J. I., Church, G. M., Fields, S., and Baker, D. (2015) A general strategy to construct small molecule biosensors in eukaryotes. *eLife* 4, No. 10606, DOI: 10.7554/eLife.10606/10.7554/eLife.10606.
- (12) Tiwari, S. B., Belachew, A., Ma, S. F., Young, M., Ade, J., Shen, Y., Marion, C. M., Holtan, H. E., Bailey, A., Stone, J. K., Edwards, L., Wallace, A. D., Canales, R. D., Adam, L., Ratcliffe, O. J., and Repetti, P. P. (2012) The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J.* 70, 855–865.
- (13) Murphy, K. F., Balazsi, G., and Collins, J. J. (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12726–12731.
- (14) Ellis, T., Wang, X., and Collins, J. J. (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* 27, 465–471.
- (15) Futcher, B., and Carbon, J. (1986) Toxic Effects of Excess Cloned Centromeres. *Mol. Cell. Biol.* 6, 2213–2222.
- (16) Tyo, K. E., Ajikumar, P. K., and Stephanopoulos, G. (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat. Biotechnol.* 27, 760–765.
- (17) Nole-Wilson, S., and Krizek, B. A. (2000) DNA binding properties of the Arabidopsis floral development protein AINTEGUMENTA. *Nucleic Acids Res.* 28, 4076–4082.
- (18) Kagaya, Y., Ohmiya, K., and Hattori, T. (1999) RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res.* 27, 470–478.
- (19) Sakuma, Y., Liu, Q., Dubouzet, J. G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2002) DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.* 290, 998–1009.
- (20) Sayou, C., Monniaux, M., Nanao, M. H., Moyroud, E., Brockington, S. F., Thevenon, E., Chahtane, H., Warthmann, N., Melkonian, M., Zhang, Y., Wong, G. K., Weigel, D., Parcy, F., and Dumas, R. (2014) A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* 343, 645–648.
- (21) Ciolkowski, I., Wanke, D., Birkenbihl, R. P., and Somssich, I. E. (2008) Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. *Plant Mol. Biol.* 68, 81–92.
- (22) Kim, J. S., Mizoi, J., Kidokoro, S., Maruyama, K., Nakajima, J., Nakashima, K., Mitsuda, N., Takiguchi, Y., Ohme-Takagi, M., Kondou, Y., Yoshizumi, T., Matsui, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012) Arabidopsis growth-regulating factor7 functions as a transcriptional repressor of abscisic acid- and osmotic stress-responsive genes, including DREB2A. *Plant Cell* 24, 3393–3405.
- (23) Kim, J. H., Choi, D., and Kende, H. (2003) The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. *Plant J.* 36, 94–104.
- (24) Wu, A., Allu, A. D., Garapati, P., Siddiqui, H., Dortay, H., Zanon, M. I., Asensi-Fabado, M. A., Munne-Bosch, S., Antonio, C., Tohge, T., Fernie, A. R., Kaufmann, K., Xue, G. P., Mueller-Roeber, B., and Balazadeh, S. (2012) JUNGBRUNNEN1, a reactive oxygen species-responsive NAC transcription factor, regulates longevity in Arabidopsis. *Plant Cell* 24, 482–506.
- (25) Matallana-Ramirez, L. P., Rauf, M., Farage-Barhom, S., Dortay, H., Xue, G. P., Droge-Laser, W., Lers, A., Balazadeh, S., and Mueller-Roeber, B. (2013) NAC transcription factor ORE1 and senescence-induced BIFUNCTIONAL NUCLEASE1 (BFN1) constitute a regulatory cascade in Arabidopsis. *Mol. Plant* 6, 1438–1452.
- (26) Allu, A. D., Brotman, Y., Xue, G. P., and Balazadeh, S. (2016) Transcription factor ANAC032 modulates JA/SA signalling in response to *Pseudomonas syringae* infection. *EMBO Rep.* 17, 1578–1589.
- (27) Garapati, P., Xue, G. P., Munne-Bosch, S., and Balazadeh, S. (2015) Transcription Factor ATAF1 in Arabidopsis Promotes Senescence by Direct Regulation of Key Chloroplast Maintenance and Senescence Transcriptional Cascades. *Plant Physiol.* 168, 1122–1139.
- (28) Christianson, J. A., Wilson, I. W., Llewellyn, D. J., and Dennis, E. S. (2009) The low-oxygen-induced NAC domain transcription factor ANAC102 affects viability of Arabidopsis seeds following low-oxygen treatment. *Plant Physiol.* 149, 1724–1738.
- (29) Yanagisawa, S. (2002) The Dof family of plant transcription factors. *Trends Plant Sci.* 7, 555–560.
- (30) Prouse, M. B., and Campbell, M. M. (2013) Interactions between the R2R3-MYB transcription factor, AtMYB61, and target DNA binding sites. *PLoS One* 8, e65132.
- (31) Jensen, M. K., Kjaersgaard, T., Nielsen, M. M., Galberg, P., Petersen, K., O'Shea, C., and Skriver, K. (2010) The Arabidopsis thaliana NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling. *Biochem. J.* 426, 183–196.
- (32) Naranjo, S., Smith, J. D., Arterio, C. G., Zhang, M., Zhou, Y., Palmer, M. E., and Fraser, H. B. (2015) Dissecting the Genetic Basis of a Complex cis-Regulatory Adaptation. *PLoS Genet.* 11, e1005751.
- (33) Gao, X., Tsang, J. C., Gaba, F., Wu, D., Lu, L., and Liu, P. (2014) Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Res.* 42, e155.
- (34) Karim, A. S., Curran, K. A., and Alper, H. S. (2013) Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Res.* 13, 107–116.
- (35) Peng, B., Williams, T. C., Henry, M., Nielsen, L. K., and Vickers, C. E. (2015) Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. *Microb. Cell Fact.* 14, 91.
- (36) Hochrein, L., Machens, F., Gremmels, J., Schulz, K., Messerschmidt, K., and Mueller-Roeber, B. (2017) AssemblX: a user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Res.*, gkx034.
- (37) Brophy, J. A., and Voigt, C. A. (2014) Principles of genetic circuit design. *Nat. Methods* 11, 508–520.
- (38) Kelwick, R., MacDonald, J. T., Webb, A. J., and Freemont, P. (2014) Developments in the tools and methodologies of synthetic biology. *Front. Bioeng. Biotechnol.* 2, 60.
- (39) Beekwilder, J., van Rossum, H. M., Koopman, F., Sonntag, F., Buchhaupt, M., Schrader, J., Hall, R. D., Bosch, D., Pronk, J. T., van Maris, A. J., and Daran, J. M. (2014) Polycistronic expression of a beta-carotene biosynthetic pathway in *Saccharomyces cerevisiae* coupled to beta-ionone production. *J. Biotechnol.* 192, 383–392.
- (40) Verwaal, R., Wang, J., Meijnen, J. P., Visser, H., Sandmann, G., van den Berg, J. A., and van Ooyen, A. J. (2007) High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive trans-

formation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. *Appl. Environ. Microbiol.* 73, 4342–4350.

(41) Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* 4, 975–986.

(42) Mitchell, L. A., Chuang, J., Agmon, N., Khunsriraksakul, C., Phillips, N. A., Cai, Y., Truong, D. M., Veerakumar, A., Wang, Y., Mayorga, M., Blomquist, P., Sadda, P., Trueheart, J., and Boeke, J. D. (2015) Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Res.* 43, 6620–6630.

(43) de Boer, C. G., and Hughes, T. R. (2012) YeTFaSCo: a database of evaluated yeast transcription factor sequence specificities. *Nucleic Acids Res.* 40, D169–179.

(44) Perez-Rodriguez, P., Riano-Pachon, D. M., Correa, L. G., Rensing, S. A., Kersten, B., and Mueller-Roeber, B. (2010) PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 38, D822–827.

(45) Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, D., Sherman, B. K., and Yu, G. L. (2000) Arabidopsis transcription factors: denome-wide comparative analysis among eukaryotes. *Science* 290, 2105–2110.

(46) Bemer, M., van Dijk, A. D., Immink, R. G., and Angenent, G. C. (2017) Cross-Family Transcription Factor Interactions: An Additional Layer of Gene Regulation. *Trends Plant Sci.* 22, 66–80.

(47) Olsen, A. N., Ernst, H. A., Leggio, L. L., and Skriver, K. (2005) NAC transcription factors: structurally distinct, functionally diverse. *Trends Plant Sci.* 10, 79–87.

(48) Shimizu-Sato, S., Huq, E., Tepperman, J. M., and Quail, P. H. (2002) A light-switchable gene promoter system. *Nat. Biotechnol.* 20, 1041–1044.

(49) Gibson, D. G. (2011) Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol.* 498, 349–361.

(50) Zhang, Y., Werling, U., and Edelman, W. (2012) SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* 40, e55.

(51) Messerschmidt, K., Hochrein, L., Dehm, D., Schulz, K., and Mueller-Roeber, B. (2016) Characterizing seamless ligation cloning extract for synthetic biological applications. *Anal. Biochem.* 509, 24–32.

(52) Gietz, R. D., and Schiestl, R. H. (2007) Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 1–4.



## **SUPPLEMENTARY INFORMATION**

### **Plant-derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces cerevisiae***

Gita Naseri<sup>1</sup>, Salma Balazadeh<sup>2,3</sup>, Fabian Machens<sup>1</sup>, Iman Kamranfar<sup>3</sup>, Katrin Messerschmidt<sup>1</sup>, Bernd Mueller-Roeber<sup>2,3,\*</sup>

<sup>1</sup>University of Potsdam, Cell2Fab Research Unit, Karl-Liebnecht-Str. 24-25, 14476 Potsdam, Germany; <sup>2</sup>Max-Planck Institute of Molecular Plant Physiology, Plant Signalling Group, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany; <sup>3</sup>University of Potsdam, Department Molecular Biology, Karl-Liebnecht-Str. 24-25, House 20, 14476 Potsdam, Germany

**\*Corresponding author:** Bernd Mueller-Roeber, [bmr@uni-potsdam.de](mailto:bmr@uni-potsdam.de)

**Supplementary Methods**

**Supplementary Figures S1 – S9**



## SUPPLEMENTARY METHODS

### General

Plasmid maps, produced with CLC Main Workbench (Qiagen, Hilden, Germany), are shown in **Supplementary Figure S9**.

### Construction of yEGFP reporter plasmid

Plasmid pFM005 was constructed by inserting the *CYC1* minimal promoter of yeast strain BY4741 between the *Bam*HI and *Eco*RI sites of pUOGB (gift from Timothy Lu, MIT). Subsequently, pFM005 was digested with *Not*I/*Eco*RI to remove *URA3* and replacing it with a partial *URA3* CDS which serves a homology donor for genome integration and contains a central *Pme*I site for vector linearization, PCR-amplified from pFM003B (see below; primers GN001/GN002), resulting in plasmid pGN005B (*MCS-Pro<sub>CYC1min</sub>-yEGFP*; **Supplementary Fig. S9**). In this vector, expression of *yEGFP* is regulated by a TF produced from an expression construct. As a positive control, we employed plasmid pGN006B, which drives *yEGFP* expression from the yeast *TDH3* promoter (*Pro<sub>TDH3</sub>-yEGFP*; **Supplementary Fig. S9**). To construct pGN006B, the *CYC1* minimal promoter of pGN005B was replaced by PCR-amplified *TDH3* promoter (primers GN003/GN004 on BY4741 genomic DNA). As a negative control, we used plasmid pGN005B, which drives *yEGFP* expression from the *CYC1* minimal promoter (*Pro<sub>CYC1min</sub>-yEGFP*). Plasmid and primer sequences are given in **Supplementary Table S2**.

### Construction of plasmids for ATF expression

Expression plasmids pGN003B (*NLS-MCS-EDLLAD*) and pGN004B (*NLS-EDLLAD-MCS*) were constructed by modifying yeast integration plasmids pFM003B and pFM004B, respectively. pFM003B (*NLS-MCS-GAL4AD*) and pFM004B (*NLS-GAL4AD-MCS*) are derivatives of pLVGI<sup>1</sup> from which the *Age*I site was deleted and the *Stu*I site replaced by *Pme*I. Furthermore, the *yEGFP* reporter was removed and replaced by the SV40\_NLS and the GAL4 AD. pFM003B contains *Bam*HI and *Age*I sites between the NLS and GAL4AD, and pFM004B contains *Bam*HI and *Age*I sites downstream of GAL4 AD. To construct pGN003B, pFM003B was digested with *Sac*II/*Bam*HI and gel-purified to remove the GAL4 AD. The remaining 6.9-kb fragment was combined with two pairs of annealed single-stranded oligonucleotides, GN005/GN006 and GN007/GN008, introducing the EDLL AD, and *Bam*HI and *Age*I sites upstream of it. pGN004B was similarly constructed with two pairs of annealed single-stranded oligonucleotides (GN009/GN010 and GN011/GN012) through which *Bam*HI and *Age*I sites were introduced downstream of EDLL AD.

To construct pGN003C (*NLS-MCS-EDLLAD-EDLLAD*), *AgeI*-digested pGN003B was assembled with annealed single-stranded oligonucleotides (GN013/GN014) introducing a second EDLL AD copy. To construct pGN005D (*MCS*), pLVGI was digested with *AgeI* and after fill-in reaction religated to obtain pLVGI\_Δ*AgeI*, which was cut with *Bam*HI/*Xho*I and gel-purified to remove the *yEGFP* reporter gene. The remaining 6.8-kb fragment was assembled with annealed single-stranded oligonucleotides (GN015/GN016) introducing *Bam*HI and *AgeI* sites. pGN003E (*NLS-MCS-Linker-EDLLAD*) is derivative of pGN003B containing 120-bp linker in front of EDLL AD. To construct pGN003E, pGN003B was digested with *Bam*HI/*AgeI* and combined with PCR-amplified EDLL AD and upstream 120-bp (primers GN017/GN018; on *Arabidopsis thaliana* Col-0 leaf cDNA). Moreover, we used pFM0031 (*MCS-NLS-VP64AD*), in which SV40\_NLS fused to VP64 AD is placed downstream of *Bam*HI and *AgeI* restriction sites. Plasmid and primer sequences are given in **Supplementary Table S3**.

### **Construction of centromeric and episomal plasmids**

To generate pGN001, pGN003B was digested with *Pme*I and *Not*I. The 6.8-kb vector backbone was assembled with the *CEN/ARS* replication origin obtained by PCR from pAG414GPD-ccdB (Addgene #14144) with primers GN019/GN020. To generate pGN002, pGN001 was cut with *Eco*RI and *Sex*AI to remove the GAL4 AD. The vector backbone was assembled with the *PGK1* promoter (primers GN021/GN022 on yeast BY4741 DNA), the *yEGFP* reporter (primers GN023/GN024 on pGN005B), and the modified yeast *GAL1* promoter carrying the *lacO* operator (primers GN025/GN026 on pGN003B). Through this, *Sbf*I and *Sfo*I sites were inserted downstream of the *PGK1* promoter, *Avr*II, *Bg*II and *Sal*I sites were inserted upstream of the *PGK1* promoter, and *Sex*AI, *Eco*RI, *Rsr*II and *Age*I sites were inserted downstream of the *GAL1* promoter. To generate pGN003, pGN002, was digested with *Sbf*I/*Sfo*I and PCR-amplified bacterial tetracycline repressor (TetR) (primers GN027/GN028 on pTVGI (1) and yeast *FBA1* terminator (primers GN029/GN030, amplified from yeast BY4741 DNA) were inserted by SLICE, resulting in pGN003. To construct pGN004, yeast *PGI1* terminator, PCR-amplified from BY4741 genomic DNA (primers GN031/GN032), a part of the modified *GAL1* promoter from pGN003B (primers GN033/GN034), and annealed single-stranded oligonucleotides GN035/GN036 were assembled with *Sex*AI/*Rsr*II-digested pGN003. Thereby, two copies of the *tetO*<sub>2</sub> locus were inserted within of the *GAL1* promoter. In the next step, the *CYC1* terminator (~400-bp) was removed from pGN004 by *Xho*I digestion. The digested plasmid was combined with *IDP1* and *TPS1* terminators (primers GN037/GN038 and GN039/GN040, respectively, on yeast BY4741 DNA) and annealed single-stranded oligonucleotides GN041/GN042 introducing the *CMV* minimal promoter, resulting in plasmid pGN005. To create pGN006, the yeast *SPG5* terminator (primers GN043/GN044; PCR on yeast BY4741 DNA) was inserted into

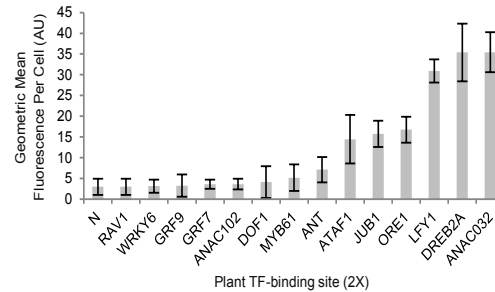
*Bsa*BI/*Xho*I-digested pGN005. In pGN006, the yeast *TEF1* promoter drives constitutive expression of the LacI inhibitor which binds to the IPTG-inducible *GAL1* promoter harboring the *lacO* operator thereby inhibiting its activity. Adding IPTG prevents binding of LacI to the operator allowing transcription of the downstream gene. pGN006 also contains an ATc-inducible *GAL1* promoter with two *tetO*<sub>2</sub> operators<sup>2</sup>. The yeast *PGK1* promoter drives constitutive expression of the TetR repressor, which binds to the *tetO*<sub>2</sub> operator sites in the modified *GAL1* promoter thereby repressing downstream gene expression; adding ATc prevents binding of TetR to the promoter leading to activation of transcription.

To construct pGN007, *yEGFP* was released from pGN006 by *Bam*HI/*Asc*I digestion and replaced with PCR-amplified *AcGFP1* CDS (primers GN045/GN046; plasmid AcGFP1-N1; Addgene #54705). Plasmid pGN008 was constructed by digesting pGN007 with *Bsa*BI and *Xho*I and inserting PCR-amplified *DsRed* (primers GN047/GN048; plasmid MSCV-CMV-DsRed-IRES-d24EGFP; Addgene #41944). To construct pGN009, pGN008 was digested with *Sph*I and *Sfo*I to release *CEN/ARS*. The plasmid backbone was then assembled with PCR-amplified *TEF1* promoter and a ~200-bp downstream fragment (primers GN049/GN050; on pGN008), PCR-amplified *2μ* origin (primers GN051/GN052; plasmid pYES2-CT; Invitrogen, V825120), and single-stranded oligonucleotides GN053/GN054, introducing *Not*I and *Sfo*I sites. Plasmid and primer sequences are given in **Supplementary Table S4**.

## REFERENCES

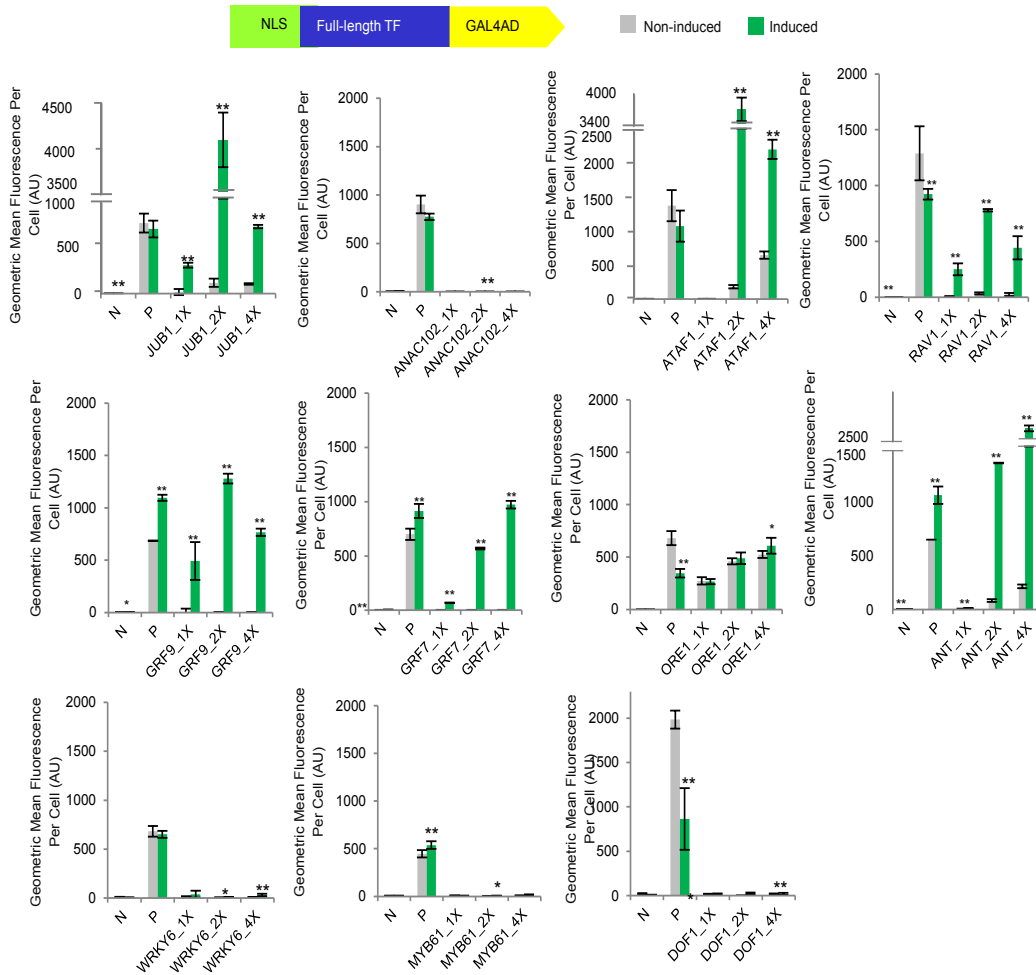
1. Ellis, T., Wang, X. and Collins, J.J. (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol*, **27**, 465-471.
2. Murphy, K.F., Balazsi, G. and Collins, J.J. (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci U S A*, **104**, 12726-12731.

## Supplementary Figure S1



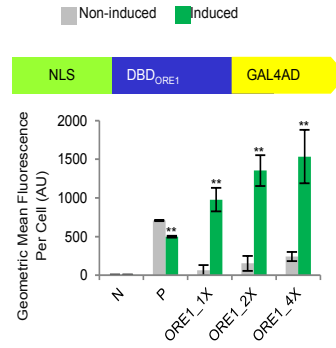
**Supplementary Figure S1.** Binding site specificity of plant TFs in yeast. Two copies ('2X') of binding sites from various plant TFs were cloned into reporter plasmid pGN005B upstream of the *CYC1* minimal promoter to drive *yEGFP* expression. The resulting plasmids were then integrated into the *ura3-52* locus of the yeast genome and fluorescence output was measured. N, empty pGN005B (*Pro<sub>CYC1min</sub>-yEGFP*). Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. AU, arbitrary units. Full data are shown in **Supplementary Data S7**.

## Supplementary Figure S2



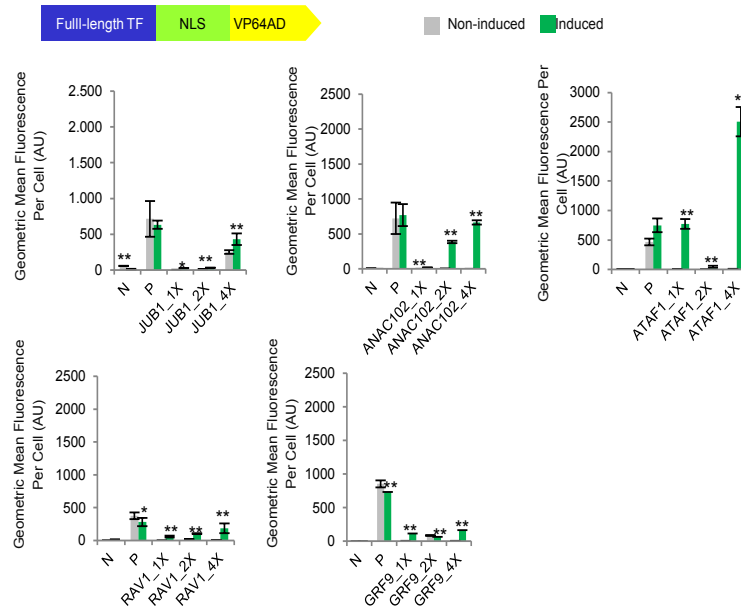
**Supplementary Figure S2.** Transcriptional output of plant-derived NLS-TF-GAL4AD regulators. The CDSs of plant TFs were inserted between an N-terminal NLS and a C-terminal GAL4 AD. The NLS-TF-GAL4AD regulators were tested for their capacity to activate yEGFP reporter gene expression from the *CYC1* minimal promoter harboring one ('1X'), two ('2X') or four ('4X') copies of the cognate binding sites. The effector and reporter constructs were chromosomally integrated into the yeast genome. yEGFP output signal was tested in the absence and presence of inducer (IPTG). Grey, non-induction medium; green, induction medium. N, negative control (*Pro<sub>CYC1min</sub>-yEGFP*); P, positive control (*Pro<sub>TDH3</sub>-yEGFP*). Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Asterisks indicate statistically significant difference from non-induction medium (Student's *t*-test; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ). AU, arbitrary units. Full data are shown in **Supplementary Data S8**.

### Supplementary Figure S3



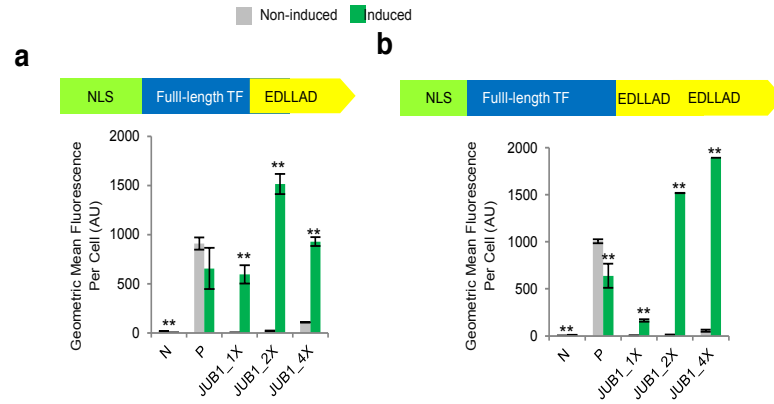
**Supplementary Figure S3.** Transcriptional output of NLS-DBD<sub>ORE1</sub>-GAL4AD. The ability of DBD<sub>ORE1</sub>-GAL4AD regulator to activate yEGFP reporter gene expression from one ('1X'), two ('2X') or four ('4X') copies of the cognate binding sites (chromosomally integrated into the yeast genome) was tested in the absence and presence of inducer (IPTG). Grey, non-induction medium; green, induction medium. N, negative control (*Pro<sub>CYC1min</sub>-yEGFP*); P, positive control (*Pro<sub>TDH3</sub>-yEGFP*). Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Asterisks indicate statistically significant difference from non-induction medium (Student's *t*-test; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ). AU, arbitrary units. Full data are shown in **Supplementary Data S9**.

## Supplementary Figure S4



**Supplementary Figure S4.** Transcriptional output of plant-derived ATFs harbouring VP64 AD. The CDSs of plant TFs were inserted upstream a NLS and the VP64 AD. The TF-NLS-VP64 regulators were tested for their capacity to activate yEGFP reporter gene expression from the *CYC1* minimal promoter harboring one ('1X'), two ('2X') or four ('4X') copies of the cognate binding sites. The effector and reporter constructs were chromosomally integrated into the yeast genome. yEGFP output signal was tested in the absence and presence of inducer (IPTG). Grey, non-induction medium; green, induction medium. N, negative control (*Pro<sub>CYC1min</sub>-yEGFP*); P, positive control (*Pro<sub>TDH3</sub>-yEGFP*). Mean fluorescence intensity per cell is given. Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Asterisks indicate statistically significant difference from non-induction medium (Student's *t*-test; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ). AU, arbitrary units. Full data are shown in **Supplementary Data S10**.

## Supplementary Figure S5

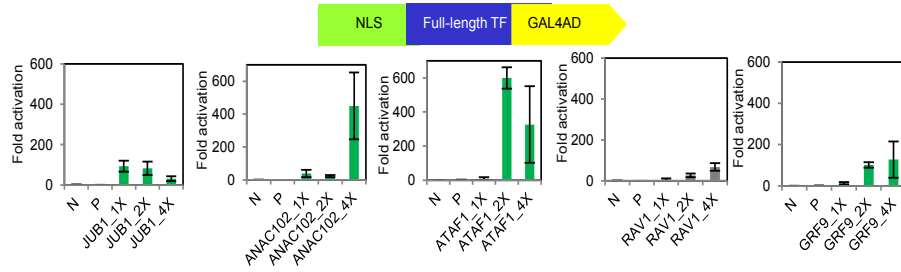


**Supplementary Figure S5.** Effect of EDLL AD on transcriptional output of JUB1. ATFs based on NAC transcription factor JUB1 were tested against one ('1X'), two ('2X') or four ('4X') copies of the JUB1 binding site placed upstream of the *CYC1* minimal promoter (chromosomal integration). **(a)** NLS-JUB1-EDLL. **(b)** NLS-JUB1-EDLL-EDLL. Grey, non-induction medium; green, induction medium. N, negative control (*Pro<sub>CYC1min</sub>-yEGFP*); P, positive control (*Pro<sub>TDH3</sub>-yEGFP*). Regulators were expressed from the IPTG-inducible *GAL1* promoter. Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Asterisks indicate statistically significant difference from non-induction medium (Student's *t*-test; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ). AU, arbitrary units. Full data are shown in **Supplementary Data S11**.

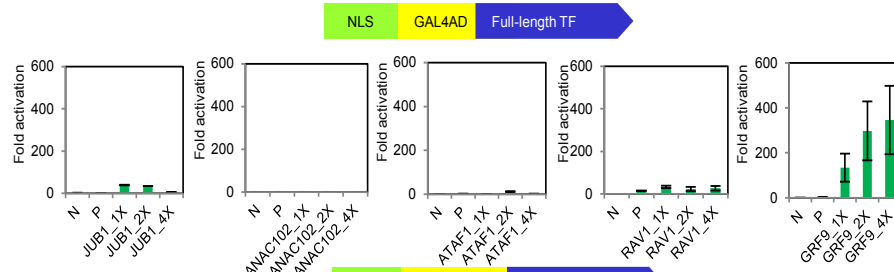


Supplementary Figure S6

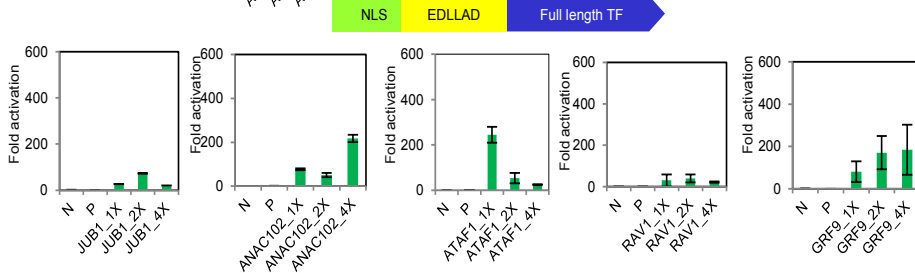
a



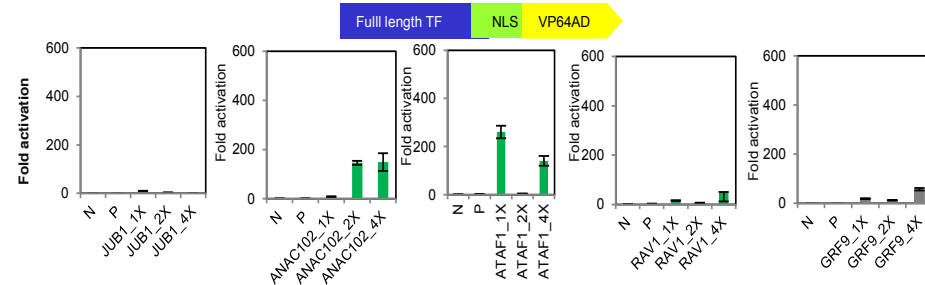
b



c

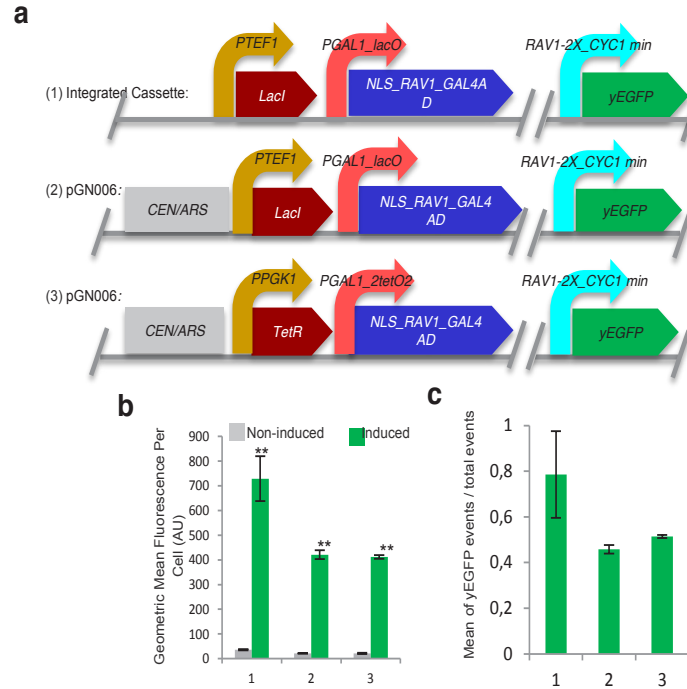


d



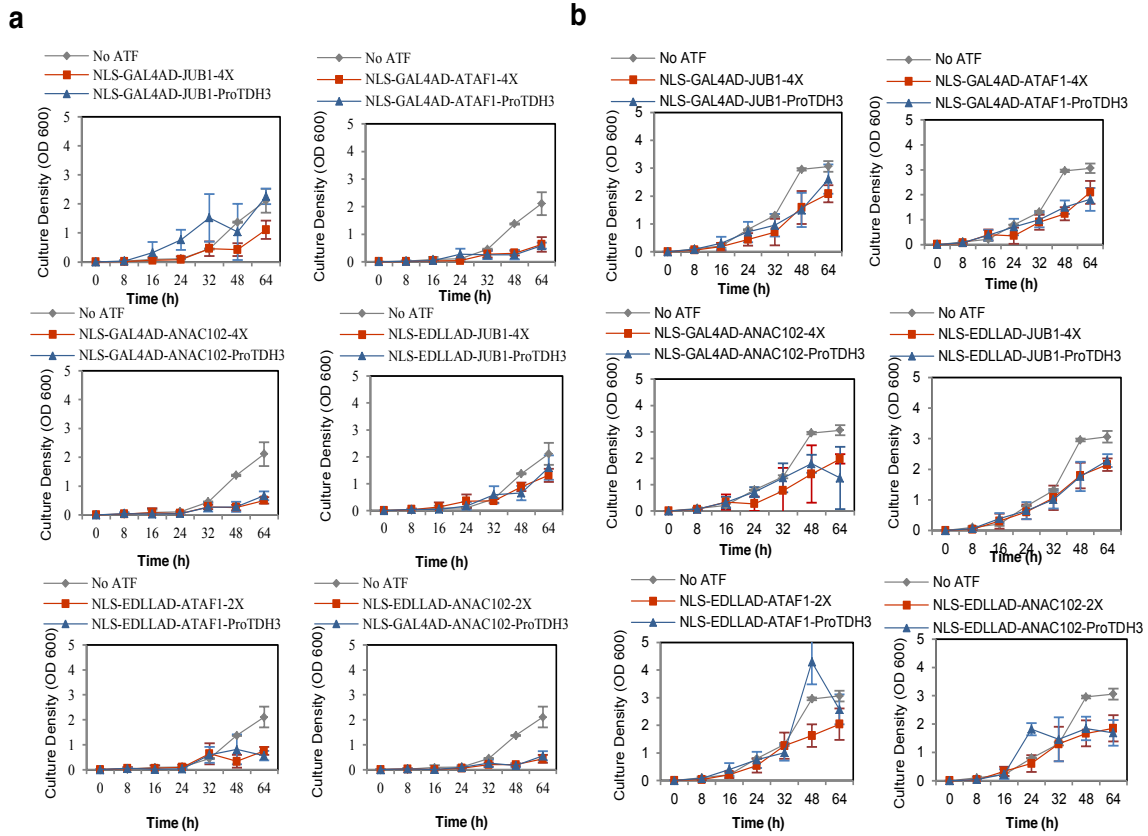
**Supplementary Figure S6.** Fold activation for plant-derived ATFs harbouring GAL4 AD, EDLL AD or VP64 AD. The activation level of five ATFs combined with (a) N- or (b) C-terminal GAL4 AD, (c) C-terminal EDLL AD, or (d) C-terminal VP64 AD compared. N, negative control (*Pro<sub>CYC1min</sub>-yEGFP*); P, positive control (*Pro<sub>TDH3</sub>-yEGFP*). Fold activation represents the ratio of yEGFP signal from induced cells to the signal from non-induced cells. Full data are shown in **Supplementary Data S12**.

## Supplementary Figure S7



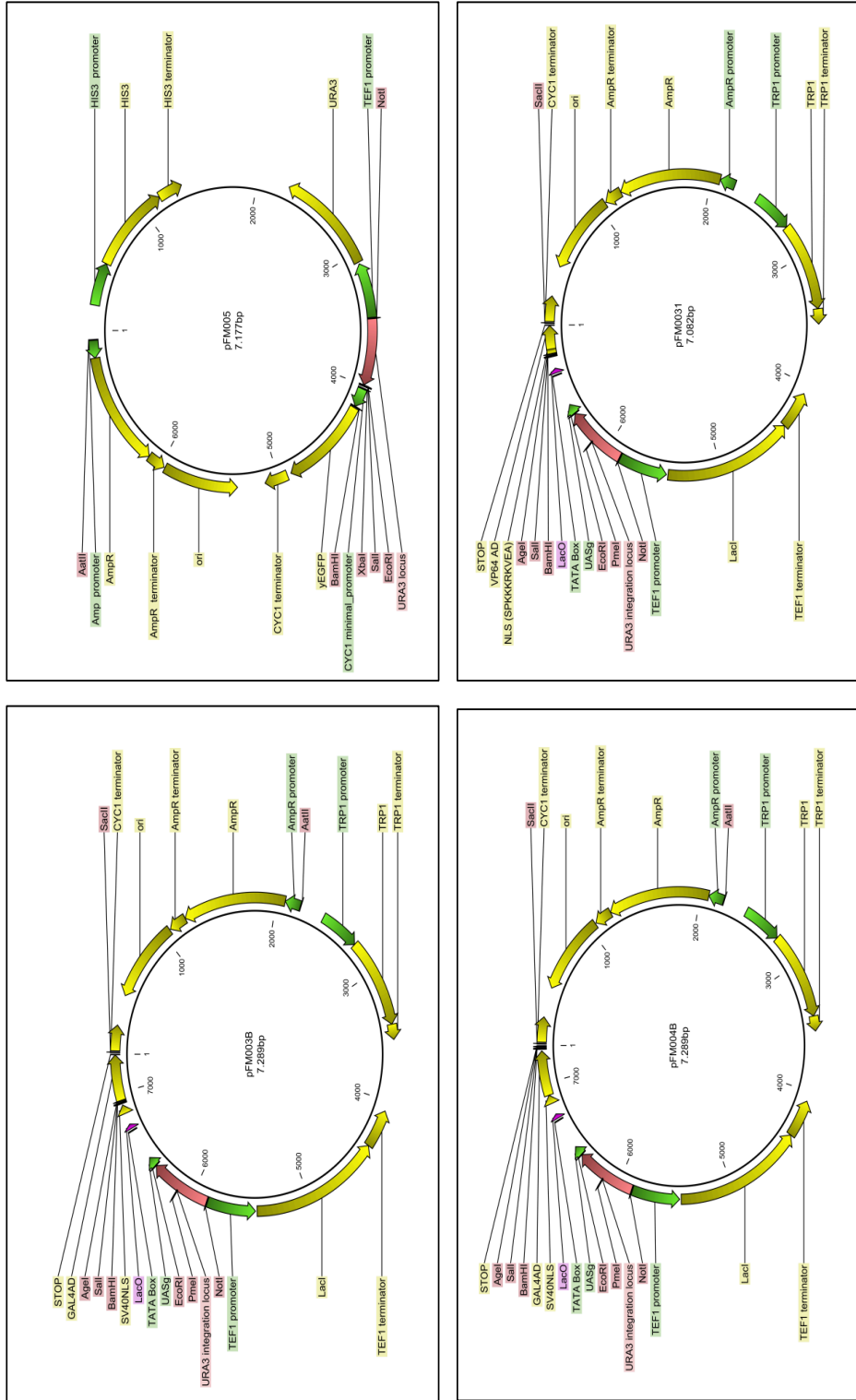
**Supplementary Figure S7.** Transcriptional output of RAV1-derived ATF integrated in yeast plasmids. **(a)** Comparison of regulatory systems implemented in the yeast genome (1), or on centromeric (*CEN/ARS*) plasmid pGN006 (2 and 3). In all cases, the NLS-RAV1-GAL4AD regulator was employed together with two RAV1 binding sites placed upstream of the *CYC1* minimal promoter (taken from pGN005B-2xBS-RAV1; **Supplementary Table S2 and S3**). In (1) and (2), NLS-RAV1-GAL4AD is expressed from the IPTG-inducible *GAL1* promoter. Expression of the ATF was induced by adding 20 mM IPTG and 2% galactose to the culture medium. In (3), NLS-RAV1-GAL4AD is expressed from the ATc-inducible *GAL1* promoter. ATF expression was induced by adding 1.08  $\mu$ M ATc or 2% galactose to the culture medium. **(b)** yEGFP output signal was tested in the absence and presence of inducer (IPTG). Grey, non-induction medium; green, induction medium. Mean fluorescence intensity per cell is given. Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Asterisks indicate statistically significant difference from non-induction medium (Student's *t*-test; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ). AU, arbitrary units. **(c)** Ratio of cells producing AcGFP1 to total cells in inducing medium. Data are mean values  $\pm$  SD obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. Full data are shown in **Supplementary Data S13**.

## Supplementary Figure S8



**Supplementary Figure S8.** Growth assays. Yeast cells with and without chromosomally integrated driver and effector constructs in **(a)** inducing medium (20 mM IPTG, 2% galactose) and **(b)** non-inducing medium (2% glucose). Grey, cells without driver and effector constructs; blue, cells with plant-derived ATFs, yEGFP expression controlled by *ProTDH3* promoter; red, cells with plant-derived ATFs controlling yEGFP expression. '2X', and '4X' indicate two or four copies of the TF binding site, respectively. Data are the mean OD<sub>600</sub> values ± SD obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates after 8, 16, 24, 36, 48 and 64 hours. Full data are shown in **Supplementary Data S14**.

Supplementary Figure S9



## **SUPPLEMENTARY INFORMATION**

### **Plant-derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces cerevisiae***

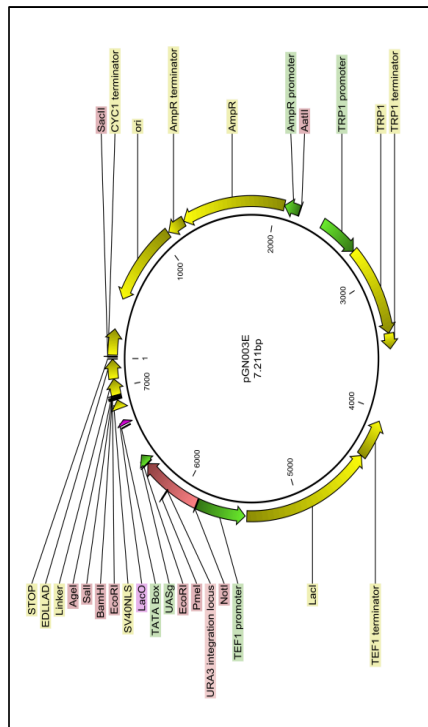
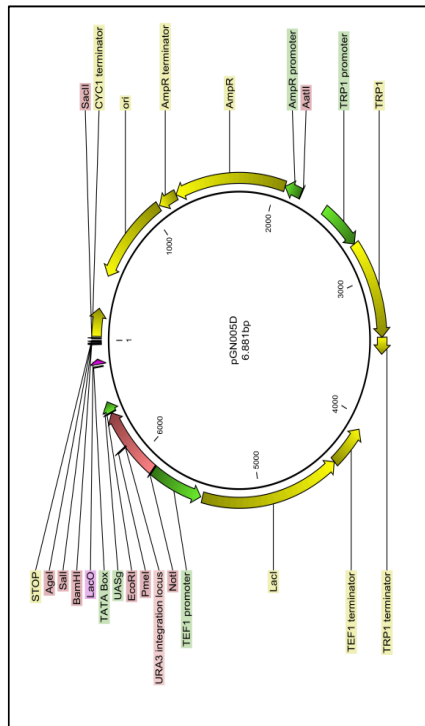
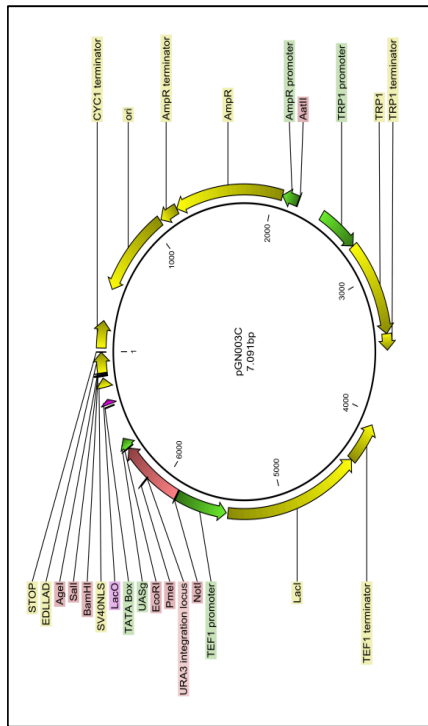
Gita Naseri<sup>1</sup>, Salma Balazadeh<sup>2,3</sup>, Fabian Machens<sup>1</sup>, Iman Kamranfar<sup>3</sup>, Katrin Messerschmidt<sup>1</sup>, Bernd Mueller-Roeber<sup>2,3,\*</sup>

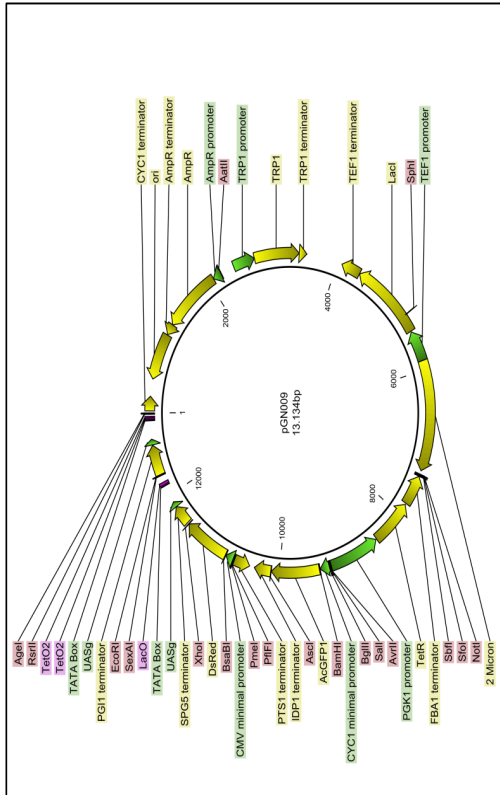
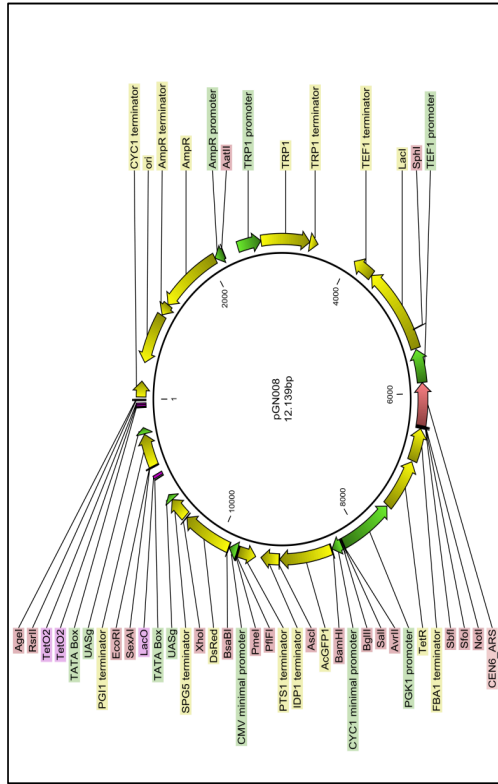
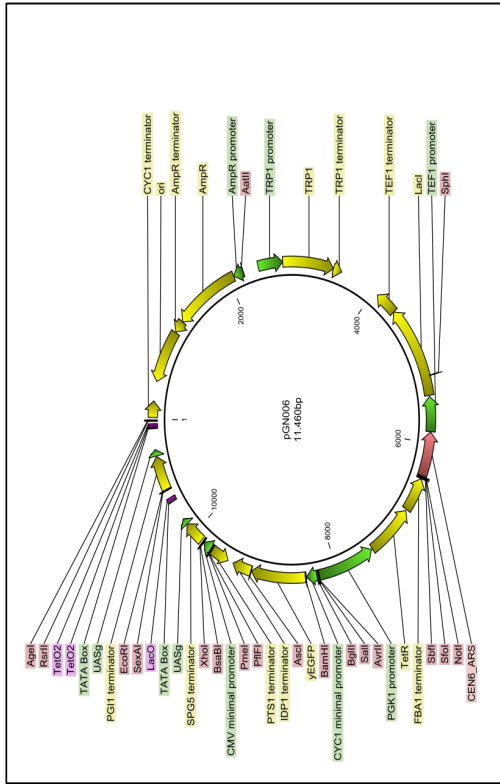
<sup>1</sup>University of Potsdam, Cell2Fab Research Unit, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany; <sup>2</sup>Max-Planck Institute of Molecular Plant Physiology, Plant Signalling Group, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany; <sup>3</sup>University of Potsdam, Department Molecular Biology, Karl-Liebknecht-Str. 24-25, House 20, 14476 Potsdam, Germany

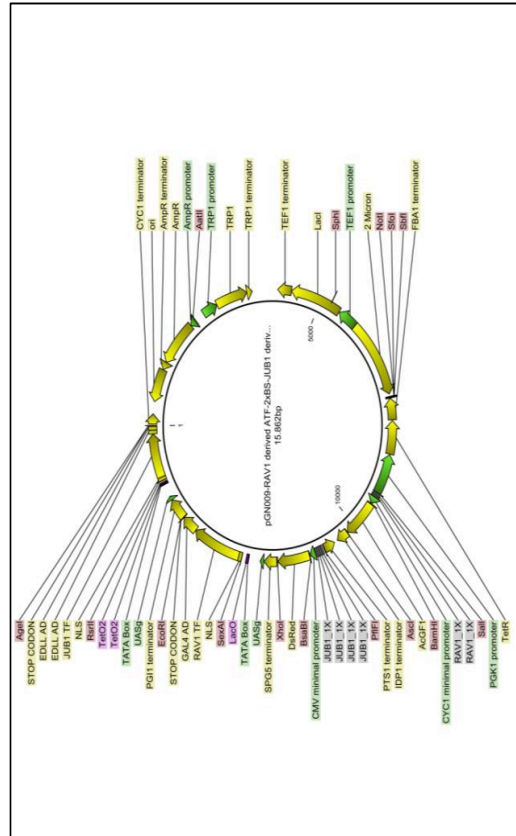
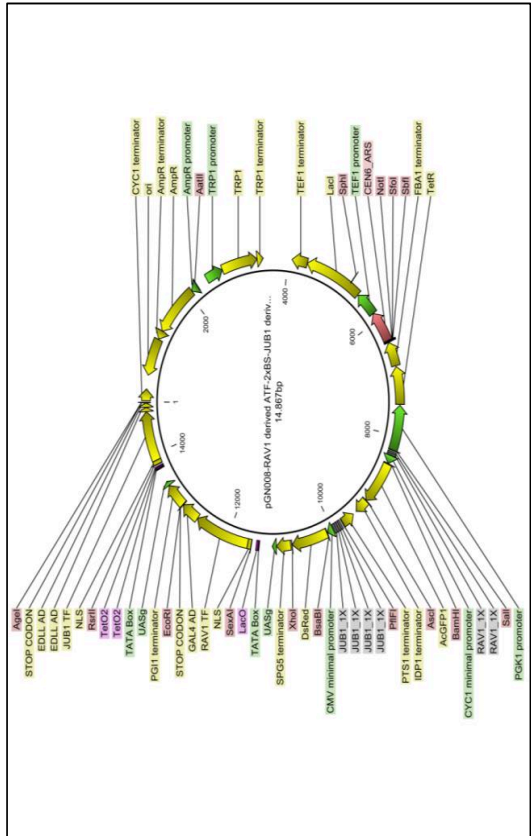
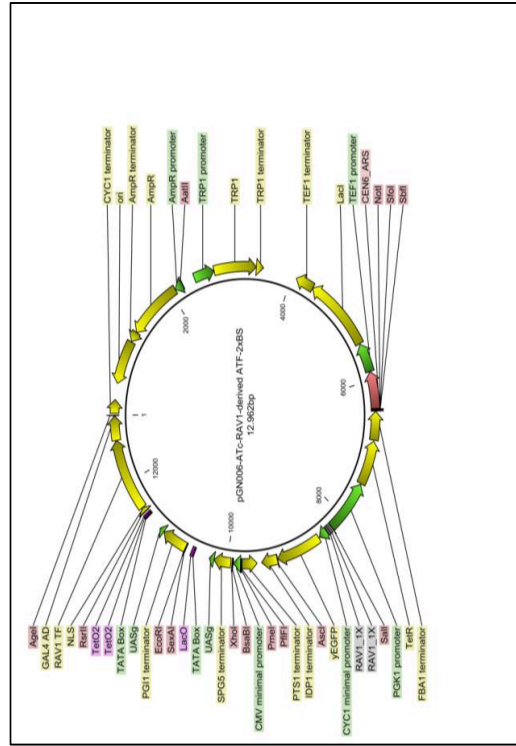
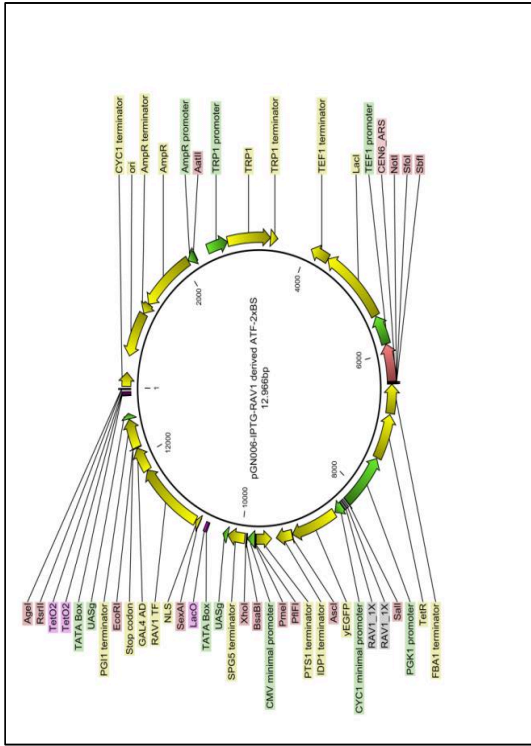
**\*Corresponding author:** Bernd Mueller-Roeber, [bmr@uni-potsdam.de](mailto:bmr@uni-potsdam.de)

**Supplementary Methods**

**Supplementary Figures S1 – S9**











## 5 COMPASS: Rapid combinatorial optimization of biochemical pathways based on artificial transcription factors

**This chapter has been published in 'Nature Communications'.**

Naseri, G. *et al.* (2019). COMPASS for rapid combinatorial optimization of biochemical pathways based on artificial transcription factors. *Nature Commun.* 10, 2615. <https://doi.org/10.1038/s41467-019-10224-x>.

Full acknowledgement is given to the other authors: Jessica Behrend, Lisa Rieper, and Bernd Mueller-Roeber and the publication source 'Nature Communications'. This is an open access article distributed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

**Abstract**

We established a high-throughput cloning method, called COMPASS for COMbinatorial Pathway ASSEMBly, for the balanced expression of multiple genes in *Saccharomyces cerevisiae*. COMPASS employs orthogonal, plant-derived artificial transcription factors (ATFs) for controlling the expression of pathway genes, and homologous recombination-based cloning for the generation of thousands of individual DNA constructs in parallel. The method relies on a positive selection of correctly assembled pathway variants from both, *in vivo* and *in vitro* cloning procedures. To decrease the turnaround time in genomic engineering, we equipped COMPASS with multi-locus CRISPR/Cas9-mediated modification capacity. In its current realization, COMPASS allows combinatorial optimization of up to ten pathway genes, each transcriptionally controlled by nine different ATFs spanning a 10-fold difference in expression strength. The application of COMPASS was demonstrated by generating cell libraries producing  $\beta$ -carotene and co-producing  $\beta$ -ionone and biosensor-responsive naringenin. COMPASS will have many applications in other synthetic biology projects that require gene expression balancing.

**Main**

The yeast *Saccharomyces cerevisiae* is a widely-used microorganism for the production of high-value chemicals<sup>1</sup>. To express heterologous enzymes, regulation at the transcriptional level is critical<sup>2, 3</sup>. To this end, several types of orthogonal artificial transcription factors (ATFs) have been made available recently<sup>4-6</sup>, including our library of 106 inducible, plant-derived ATFs of varying strengths<sup>7</sup>. Importantly, high-level expression of pathway genes often increases metabolic burden leading to growth inhibition<sup>8-13</sup>. To overcome this, methods for balancing metabolic flux have recently been developed, including CRISPR-AID, a tri-functional CRISPR system for combinatorial pathway optimization in yeast<sup>14</sup>. However, CRISPR-AID and other combinatorial optimization approaches still rely on the constitutive and strong expression of pathway genes<sup>14, 15</sup> likely affecting metabolic performance of the cell<sup>16</sup>, while inducible ATFs allow a conditional expression of genes<sup>17, 18</sup>. Hence, in COMPASS, the isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG)-inducible *GAL1* promoter controls the expression of plant-derived regulators<sup>19</sup>, and these regulators then control the expression of pathway genes through their respective binding sites implemented in the upstream promoters.

COMPASS represents a unique, high-throughput gene assembly method; it employs *in vivo* and *in vitro* homologous recombination to build large DNA constructs while at the same time eliminating unwanted scar sequences<sup>20</sup>. COMPASS makes use of positive selection protocols, which severely reduces the need to checking individual constructs and strongly improves the

efficiency of detecting correct assemblies in cloning reactions.

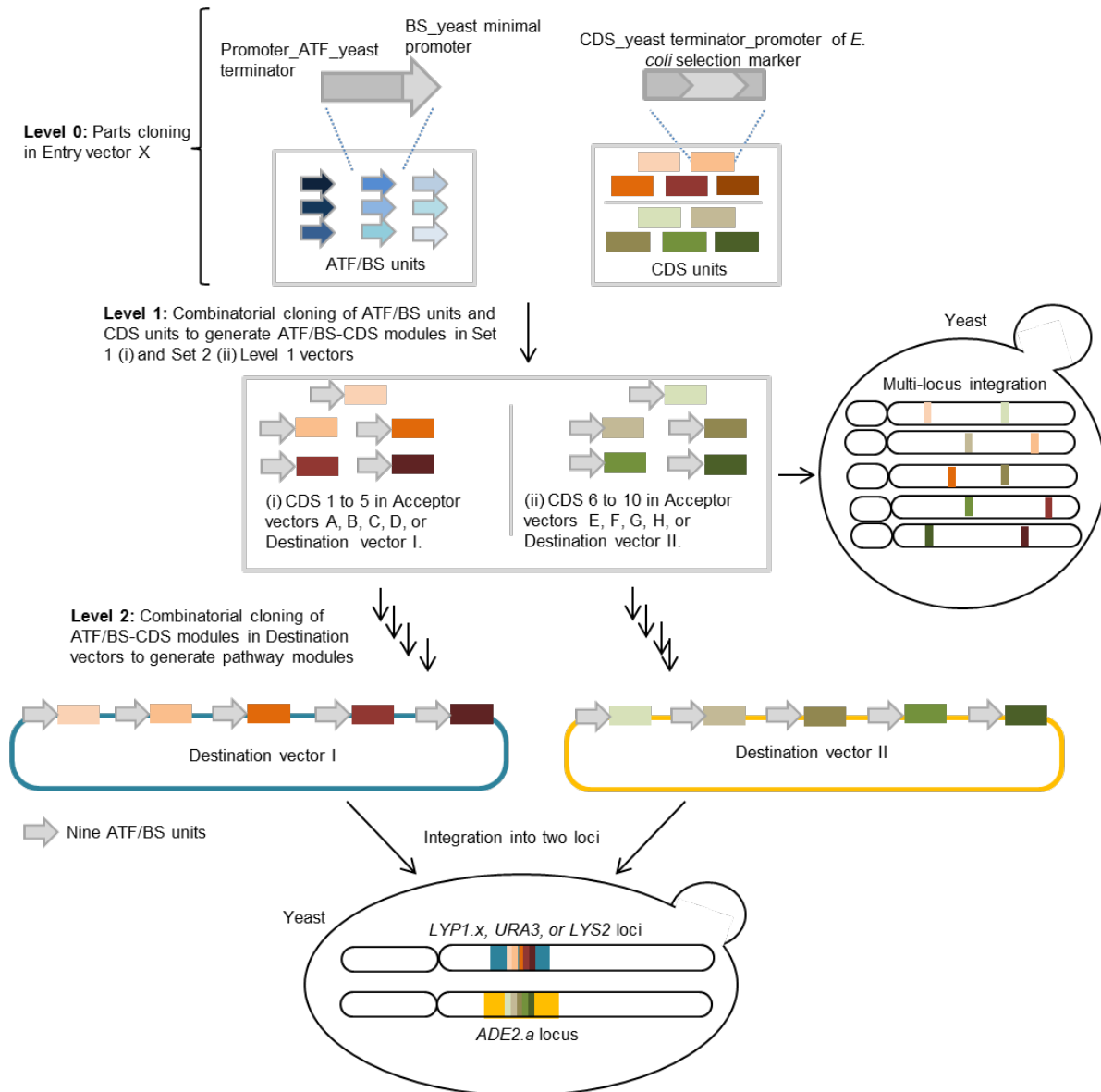
Plasmid-based systems are commonly used for pathway engineering because of the ease of their manipulation, although they often lack sufficient robustness due to segregational and structural instability<sup>21</sup>. Therefore, we adopted COMPASS for both, plasmid- and genomic integration-based pathway assemblies.

To establish complex construct libraries leading to high levels of metabolic product output, the level of the wanted product must be screenable. The diversity of a library can e.g. be assessed by screening individual colonies for the production of colored products<sup>22</sup>, such as  $\beta$ -carotene<sup>23</sup>. However, most chemicals are uncolored and their detection therefore requires alternative methods. Of note, several high-throughput screening methods have recently been developed for the detection of chemicals, e.g. a biosensor for the detection of naringenin (NG) in single yeast cells<sup>24</sup>, which we employ here within the framework of our construction method. Collectively, COMPASS implements multiple features as a new high-throughput combinatorial cloning tool for yeast synthetic biology applications.

## Results

### General outline of the combinatorial COMPASS strategy

COMPASS allows the rapid combinatorial assembly of up to ten pathway genes, each transcriptionally controlled by nine inducible ATF/binding site (BS) units differing in expression strength. In COMPASS, construct libraries are built by passing through three successive cloning levels (**Fig. 1**). At Level 0, (i) ATF/BS units and (ii) CDS units (consisting of the enzyme coding sequence (CDS), a yeast terminator, and an *Escherichia coli* promoter of a selection marker gene) are constructed in first week. Level 1 serves the combinatorial assembly of ATF/BS units upstream of up to ten CDS units to generate complete ATF/BS-CDS modules in one week. Finally, at Level 2, up to five ATF/BS-CDS modules are combinatorially assembled into a single vector in four weeks. At Level 1 and 2, successful assemblies are selected by plating cells on appropriate selection media. Level 0, 1 and 2 allow multiple parallel assemblies. Modules in Levels 1 and 2 can be integrated into the genome to generate stable yeast strains, facilitated by the CRISPR/Cas9-mediated modification system that allows one-step integration of multiple groups of cassettes into multiple loci<sup>25</sup>. Details of our approach will be described in the following.



**Fig. 1. The COMPASS workflow.**

The workflow encompasses three assembly levels. At Level 0, nine ATF/BS units are assembled into Entry vector X. Each ATF/BS unit harbors an inducible promoter to drive ATF expression, the ATF CDS, a yeast terminator, and promoter fragments containing one or more copy(s) of the ATF's BS within a minimal *CYC1* promoter. In separate reactions, each of the ten CDSs is combined with a yeast terminator and the promoter of an *E. coli* selection marker (which defines the vector of the next assembly level) into Entry vector X to generate CDS units. At Level 1, two parallel combinatorial clonings are performed: (i) The nine different ATF/BS units, five CDS units, and a set of linearized auxotrophic marker vectors (Acceptor vectors A, B, C and D, and Destination vector I) are subjected to overlap-based cloning in a single reaction tube (using NEBuilder HiFi). (ii) Similarly, nine ATF/BSs, five other CDSs, and a set of dominant-marker vectors (Acceptor vectors E, F, G and H, and Destination vector II) are combined. Using this approach, COMPASS generates ten groups of vectors, each containing nine plasmids (90 plasmids in

total). The ten groups of ATF/BS-CDS modules are integrated into ten defined loci of the genome. At Level 2, ten libraries of ATF/BS-CDS modules are established in Destination vector I and Destination vector II (five libraries in each vector). The Destination vectors I are integrated into the *LYS2*, *URA3*, (or *LYP1.x*) loci, while Destination vectors II are integrated into the *ADE2.a* locus.

### Plant-derived ATF/BS units

To construct orthogonal plant-based transcription regulatory units, we selected nine ATF/BS combinations from our previously reported library of 106 genome-integrated ATF/BSs to cover weak (NLS-GAL4AD-RAV1/4X, NLS-DBD<sub>JUB1</sub>-GAL4AD/2X, and ANAC102-NLS-VP64AD/4X; 300 – 700 arbitrary units (AU), determined using yEGFP as reporter<sup>7</sup>), medium (NLS-GAL4AD-GRF7/4X, NLS-GAL4AD-ANAC102/4X and NLS-JUB1-EDLLAD-EDLLAD/4X; 1,100 - 1,900 AU) and strong (NLS-GAL4AD-ATAF1/2X, NLS-ATAF1-GAL4AD/2X and NLS-JUB1-GAL4AD/2X; 2,500 - 4,000 AU) transcriptional outputs (**Supplementary Fig. 1**).

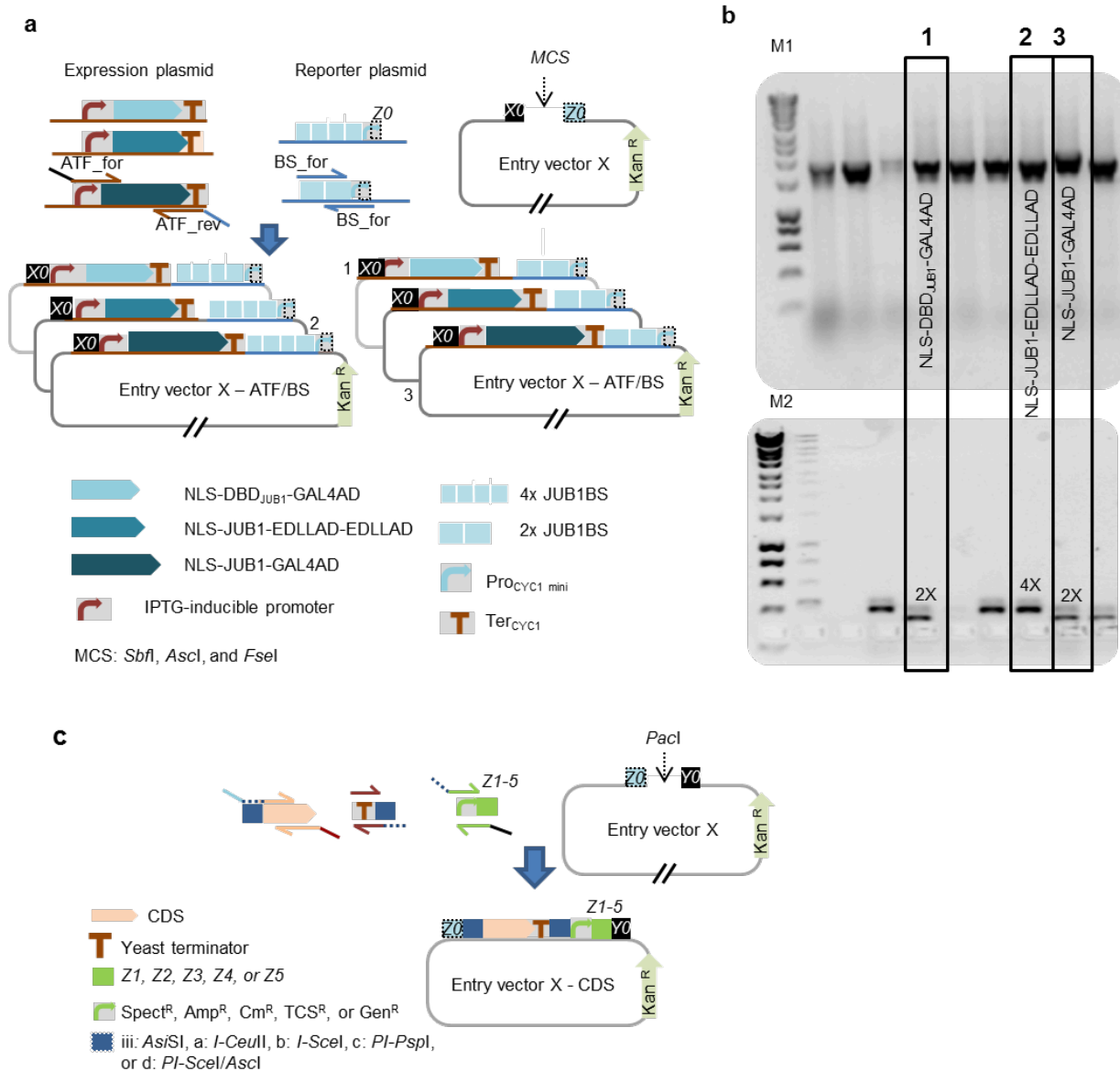
### COMPASS vectors

To optimize pathway generation using COMPASS, we designed Entry vector X as well as Acceptor and Destination vectors (see **Supplementary Note 1** and **Supplementary Table 1**). Entry vector X (**Supplementary Fig. 2**) is used to assemble (i) ATF/BS units and (ii) CDS units. Acceptor and Destination vectors are designed in two Sets and are used to assemble (i) the library of ATF/BS units upstream of CDS units at Level 1 and (ii) the library of ATF/BS-CDS modules at Level 2. Set 1 includes Destination vector I (**Supplementary Fig. 3a**) and Acceptor vectors A, B, C, and D (**Supplementary Fig. 3b**), while Set 2 includes Destination vector II (**Supplementary Fig. 4a**) and Acceptor vectors E, F, G, and H (**Supplementary Fig. 4b**). Destination vector I can be integrated into the *URA3* or *LYS2* locus of yeast, while Destination vector II can be integrated into the *ADE2.a* locus<sup>25</sup>. Additionally, Destination vector I.1 (**Supplementary Fig. 5**) allows integration into the *LYP1.x* locus<sup>25</sup>.

### Level 0:

*Construction of the ATF/BS library:* We employed combinatorial cloning (see **Supplementary Protocol**) to establish the nine combinations of ATFs and BSs in Entry vector X. For example, to build NLS-DBD<sub>JUB1</sub>-GAL4AD/2X, NLS-JUB1-EDLLAD-EDLLAD/4X and NLS-JUB1-GAL4AD/2X-ATF units, triplex PCR amplified the *Pro<sub>mGAL1-LacI</sub>-JUB1*-derived ATF fragments (from expression plasmids<sup>7</sup>) and duplex PCR amplified *Pro<sub>CYC1</sub>* containing two and four copies of the BS fragments (from reporter plasmids<sup>7</sup>) were used for overlap-based recombinational cloning

(**Supplementary Protocol**). Overlap-based recombination theoretically leads to six different ATF/BS combinations (**Fig. 2a**), of which three (number 1, 2, and 3) were needed here. Colony-PCR (**Fig. 2b**), followed by sequencing, was performed to identify the respective combinations of ATFs and BSs.



**Fig. 2. Combinatorial assembly of ATF/BS or CDS units in Entry vector X.**

(a) Combinatorial insertion of JUB1-derived ATF/BS units in Entry vector X. Three-partite fragments, harboring an IPTG-inducible promoter, the ATF, and the *CYC1* terminator are PCR-amplified from expression plasmids NLS-DBD<sub>JUB1</sub>-GAL4AD, NLS-JUB1-EDLLAD-EDLLAD, and NLS-JUB1-GAL4AD<sup>7</sup>. Two synthetic promoters, harboring two and four copies of the JUB1BS, are PCR-amplified from reporter plasmids<sup>7</sup>. ATFs and BSs amplicons are mixed in 1:10 molar ratio and cloned into *Fse*I/*Asc*I-linearized Entry vector X. The diagram shows six possible outcomes. The three required JUB1-derived

ATF/BSs, NLS-DBD<sub>JUB1</sub>-GAL4AD/2X, NLS-JUB1-EDLLAD-EDLLAD/4X, and NLS-JUB1-GAL4AD/2X, are numbered 1, 2, and 3, respectively. For simplicity, the IPTG-inducible promoter and terminators are not included in the figure. **(b)** Gel electrophoresis to identify required JUB1-derived ATF/BSs. The verification of constructs number 1, 2, and 3 was done using colony PCR on ATF/BS fragments. M1: HyperLadder 1 kb (Bioline). M2: HyperLadder V (Bioline). **(c)** Insertion of CDS units into Entry vector X. CDSs with rare RE sites (iii, a, b, c, d) in their 5' regions, a yeast terminator carrying similar RE sites in the 3' region, and promoters of *E. coli* markers fused to Z1, Z2, Z3, Z4, or Z5 are cloned into *PacI*-digested Entry vector X. The Z and RE sites define based on Level 1 vectors (Destination vectors I/II: Z1 and iii; Acceptor vectors A/E: Z2 and a; B/F: Z3 and b; C/G: Z4 and c; and E/F: Z5 and d). The HRs X0, Z0 - Z5, and Y0 are explained in footnote to **Supplementary Table 2**.

*Units of pathway genes:* The CDS, a yeast terminator and a promoter of an *E. coli* selection marker are assembled in *PacI*-digested Entry vector X (**Fig. 2c, Supplementary Protocol and Supplementary Table 2**). Introducing rare RE cleavage sites in the CDS units (see the protocol for primer design in **Supplementary Protocol**) allows replacing the CDS-terminators with other CDS-terminators in COMPASS vectors at a later stage, if needed.

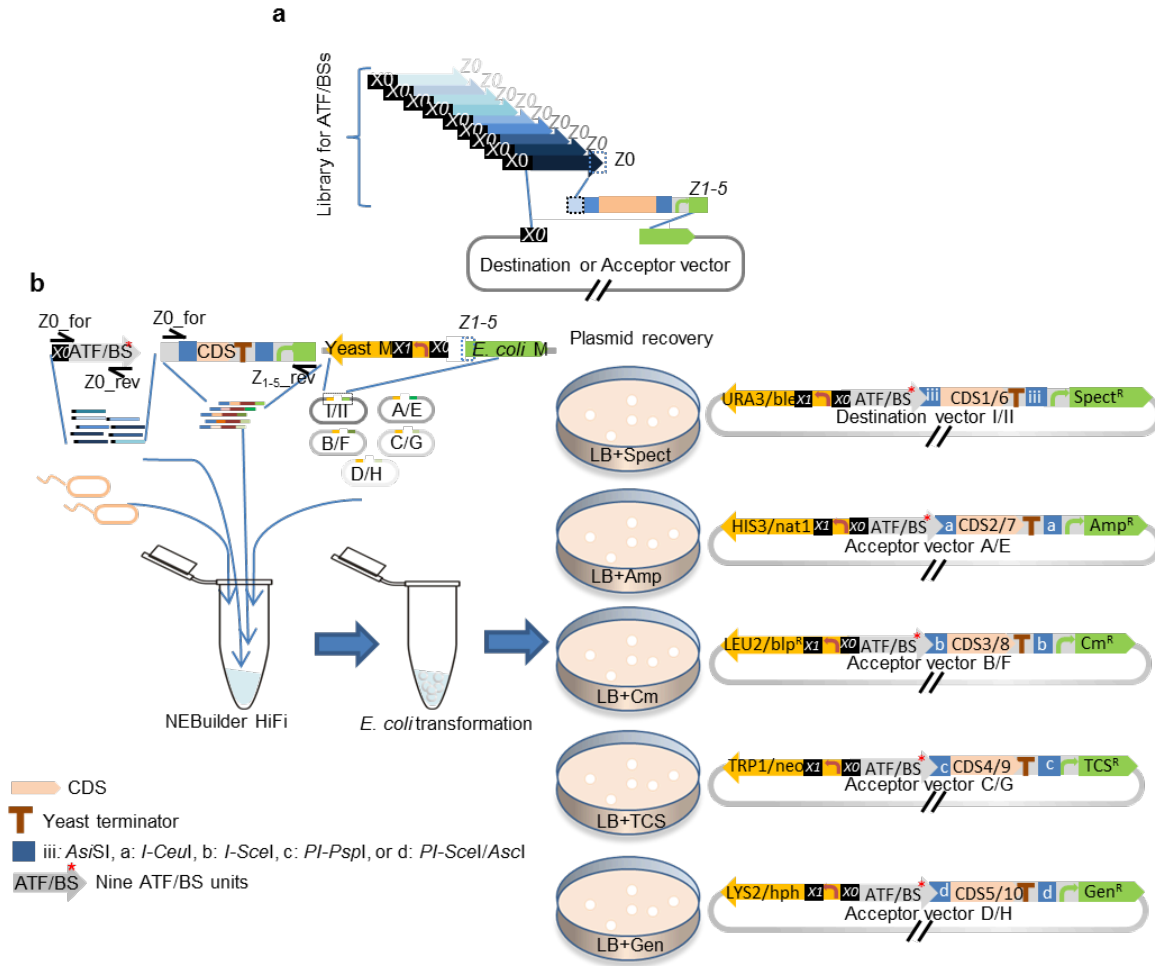
#### Level 1:

##### Construction of ATF/BS-CDS modules

To construct the library of the ATF/BS-CDS modules (**Fig. 3a**), (i) nine ATF/BS units are combinatorially assembled upstream of five CDS units in Set 1 vectors and (ii) five other CDS units are assembled in Set 2 vectors. To conduct five clonings in Set 1 vectors simultaneously (**Fig. 3b**), equal amounts of the nine ATF/BS units, five freely selected CDS units, and linearized Set 1 vectors are mixed in a single reaction tube to perform *in vitro* overlap-based cloning. By providing the missing promoter of *E. coli* selection markers within the assembly units, successful assemblies are identified by plating *E. coli* cells transformed with the reaction cocktail on media containing antibiotics. This creates a complexity of  $9 \times 5 = 45$  different plasmids in a successful experiment. Step-by-step protocols and information about primer design are given in **Supplementary Protocol**. Using the same strategy five other CDS units can be assembled in Set 2 vectors.

Moreover, CDS units can be generated through combinatorial assembly of five CDSs, five yeast terminators and promoters of five antibiotic resistance genes in either Set 1 or Set 2 vectors (see **Supplementary Protocol and Supplementary Fig. 6**).





**Fig. 3. Combinatorial assembly of ATF/BS and gene units in the modules.**

**(a)** Combinatorial assembly of ATF/BS units upstream of CDS units. Nine ATF/BS units and CDS units from the Entry vectors X are PCR-amplified. The 5' regions of ATF/BS units overlap with the X0 sequence of the vector backbone, while their 3' regions are identical to the 30 bp of the 3' end of the minimal *CYC1* promoter (Z0) and overlap with the forward primer amplifying the CDS units. The 3' regions of the CDS units overlap with Z1 - Z5 of five linearized vectors. **(b)** The COMPASS workflow to generate ATF/BS-CDS modules in vectors of either Set 1 or Set 2 (1/2). Equal amounts of the PCR-amplified ATF/BS unites (primers X0\_for/Z0\_rev, on Entry vectors X-ATF/BS), five CDS units (primers Z0\_for/Z1\_rev/Z2\_rev/Z3\_rev/Z4\_rev/Z5\_rev at concentration ratio 5:1:1:1:1:1, on mixed Entry vectors X-CDS) and linearized Destination vectors I/II and Acceptor vectors A/E, B/F, C/G, and D/H are mixed for *in vitro* overlap-based cloning in a single tube to generate different ATF/BS-CDS modules in the diverse vectors by providing the missing promoter sequences of the *E. coli* markers within the assembly units. Therefore, libraries of Destination vectors I-CDS1/II-CDS6, Acceptor vectors A-CDS2/E-CDS7, B-CDS3/F-CDS8, C-CDS4/G-CDS9, or D-CDS5/H-CDS10 are generated. X0\_for overlaps with X0, Z0\_rev overlaps with Z0\_for, while primers Z1\_rev - Z5\_rev overlap with the downstream (right) HR of the linearized vector

(called Z1 - Z5). For simplicity, IPTG-inducible promoters and terminators are not included in the figure. X0, Z0 - Z5 are explained in footnote to **Supplementary Table 2**.

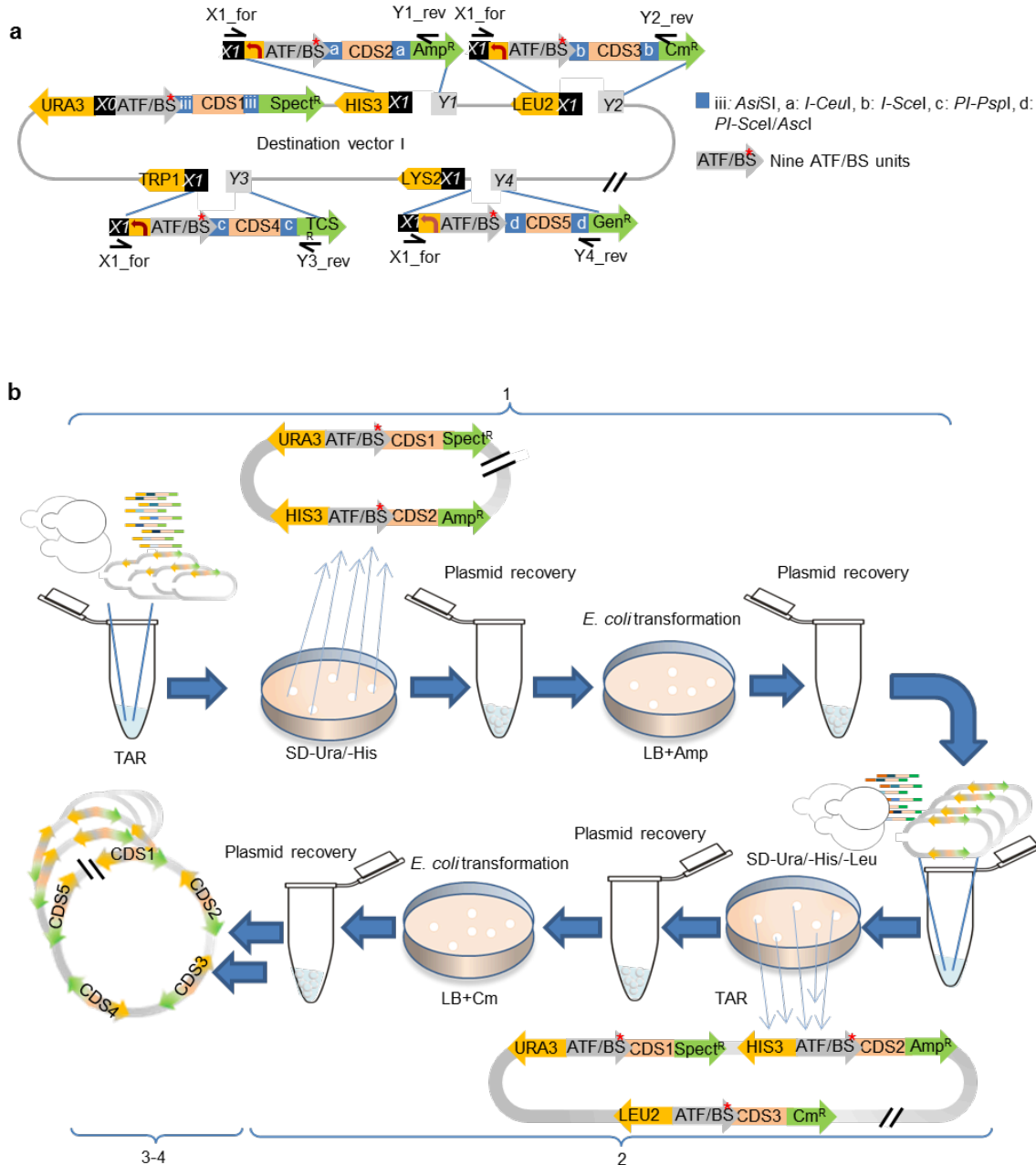
### General strategy for multi-locus integration of gene modules

We implemented the CRISPR/Cas9 strategy for multi-locus integration into the previously characterized integration sites shown to exhibit high integration efficiency to decrease the turnaround time in metabolic engineering projects. Multiple groups of donors are integrated at multiple loci, whereby each group is integrated into a single locus of different yeast cells. Each donor contains a yeast selection marker, an inducible ATF/BS, a CDS, and a yeast terminator amplified from Level 1 vectors, in addition to 50-bp up- and down-stream homology regions (HRs) allowing integration into pre-designed genomic loci. The resulting library of yeast strains grows on plates containing appropriate yeast selection markers.

### Level 2: Assembly of the pathway library

COMPASS is designed for multi-step combinatorial cloning of the multi-libraries of CDS modules ( $Pro_{\text{yeast\_auxotrophic/dominant\_marker}}-Pro_{mGAL1}-ATF/BSs-CDS-Ter_{\text{yeast}}-Pro_{E.coli\_selection\_marker}-CDS_{E.coli\_selection\_marker}-Ter_{E.coli}$ ) into the Destination vectors (library of one CDS modules in each step), as decreasing the number of inserts in an assembly reaction increases cloning efficiency<sup>26</sup>. Our approach is based on the positive selection of successful constructs from both, *in vitro* and *in vivo* cloning procedures. TAR is the preferred method over methodologies that make use of *E. coli*, because some assembly products are not stable or clonable in *E. coli*<sup>27</sup>. COMPASS allows ATF/BS to control the expression of up to five CDSs in Destination vector I (**Fig. 4a**), and up to five CDSs in Destination vector II (**Supplementary Fig. 7a**). Integration of the ten CDS modules occurs at sites *p1* (CDS1 and CDS6), *p2* (CDS2 and CDS7), *p3* (CDS3 and CDS8), *p4* (CDS4 and CDS9), *p5* (CDS5) and *p6* (CDS10) (**Supplementary Fig. 3 and 4**). As shown in **Fig. 4b** (and described in **Supplementary Protocol**), nine ATF/BS-CDS2 modules are assembled at site *p2* of nine Destination vectors I-CDS1 to generate a library of 81 Destination vectors I-CDS1-CDS2. Cells with successful constructs grow on SC-Ura/His medium, as the backbone and insert carry the *URA3* and *HIS3* selection markers, respectively. The plasmid library is recovered from the yeast cells, transformed into *E. coli*, and grown on LB agar plates containing ampicillin, where only cells harboring correct assemblies will grow. In three rounds of combinatorial cloning, ATF/BS-CDS3, ATF/BS-CDS4, and ATF/BS-CDS5 modules are successively assembled in sites *p3*, *p4*, and *p5*, respectively, of Destination vectors I-CDS1-CDS2 to generate the complete Destination vectors I-CDS1-CDS2-CDS3-CDS4-CDS5

library. The number of constructs (library size) after each cloning step is  $X_n = Y^n$ , with  $n =$  cloning step (1, 2, 3, 4, or 5),  $Y =$  number of regulators. Using the same pair of primers (**Supplementary Protocol**), libraries of ATF/BS-CDS7, ATF/BS-CDS8, ATF/BS-CDS9, ATF/BS-CDS10 modules are assembled successively at sites  $p_2$ ,  $p_3$ ,  $p_4$ , and  $p_6$ , respectively, starting with the Destination vectors II-CDS6 library (**Supplementary Protocol, Supplementary Fig. 7b**).



**Fig. 4. Combinatorial cloning of pathway modules into Destination vector I.**

**(a)** Combinatorial cloning of pathway genes into Destination vector I. The libraries of PCR-amplified (i)  $Pro_{HIS3}\text{-ATF/BS-CDS2-}Pro_{Amp^R}\text{-}Amp^R\text{-}Ter_{Amp^R}$  (primers X1\_for/Y1\_rev, on Acceptor vector A), (ii)  $Pro_{LEU2}\text{-ATF/BS-CDS3-}Pro_{Cm^R}\text{-}Cm^R\text{-}Ter_{Cm^R}$  (primers X1\_for/Y2\_rev, on Acceptor vector B), (iii)  $Pro_{TRP1}\text{-ATF/BS-}$

*CDS4-Pro<sub>TCSR</sub>-TCSR-Ter<sub>TCSR</sub>* (primers X1\_for/Y3\_rev, on Acceptor vector C), and (iv) *Pro<sub>LYS2</sub>-ATF/BS-CDS5-Pro<sub>GenR</sub>-GenR-Ter<sub>GenR</sub>* (primers X1\_for/Y4\_rev, on Acceptor vector D) modules are successively assembled in sites *p2*, *p3*, *p4*, and *p5*, respectively, starting with the Destination vectors I-CDS1 library (see **Fig. 3**), in four rounds of combinatorial cloning. **(b)** The COMPASS workflow for combinatorial assembly of ATF/BS and gene modules into Destination vector I. The mixed ATF/BS-CDS1 modules are assembled using TAR in site *p2* of Destination vectors I-CDS1. Yeast cells with successful constructs grow on SC-Ura/-His medium. Cells are scraped from the plates, the plasmid library is extracted to obtain a pool of all randomized members, transformed into *E. coli*, and cells are grown on LB plates containing ampicillin. Cells are scraped from the plates to extract the plasmid library (1). The ATF/BS-CDS3 modules are assembled in site *p3* of the Destination vectors I-CDS1-CDS2 library. Yeast cells with successful constructs grow on SC-Ura/-His/-Leu medium (2). The libraries of ATF/BS-CDS4 and ATF/BS-CDS5 modules are cloned into sites *p4* and *p5*, respectively (3 - 4). For simplicity, the IPTG-inducible promoters and terminators are not included in the figure. X0 and X1 are explained in footnote to **Supplementary Table 2**. Y1 - Y4 overlap with the last 30 bp of terminators of the *Amp<sup>R</sup>*, *Cmr<sup>R</sup>*, *TCS<sup>R</sup>*, and *Gen<sup>R</sup>* genes.

### Single-locus integration of pathways assembled in Destination vectors

We studied the production of  $\beta$ -carotene by applying the following three approaches: (1) integration of three modules into three loci of the genome; (2) transformation of Destination vector I containing the pathway modules into the cell without genome integration; and (3) integration of the Destination vector I into the genome. The strong yeast *TDH3* promoter<sup>31</sup> upstream of all three  $\beta$ -carotene CDSs served as positive control in each case. The combination of ATF/BS fragments in three colonies of each approach showing the most intense  $\beta$ -carotene accumulation was identified, and one of them was analyzed by HPLC. We observed that approach 2 (**Supplementary Fig. 9**), 1 (**Fig. 5**), and 3 (**Supplementary Fig. 10**) resulted in the highest to lowest amount of  $\beta$ -carotene accumulation.

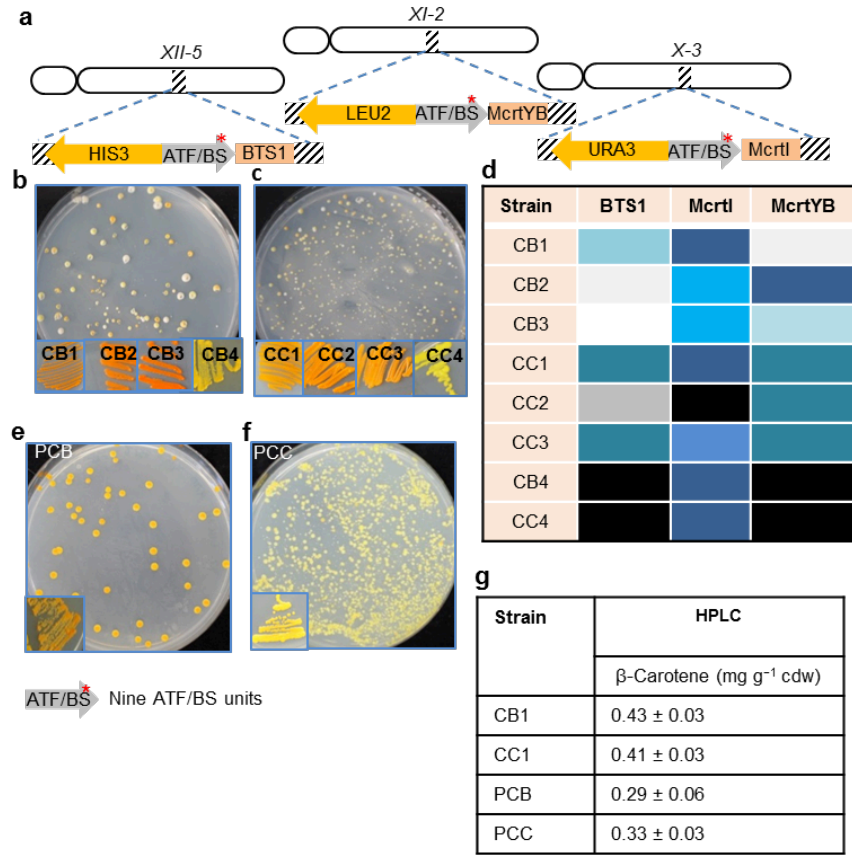
In approach 1 (**Fig. 5a**, see **Online Methods**), a wide spectrum of colors ranging from light yellow to deep orange was observed in the different colonies obtained (Gen 0.1, **Fig. 5b**; IMX672.1, **Fig. 5c**). Sequencing ATF/BS fragments (**Fig. 5d**) revealed that weak/medium, medium/strong, and medium/strong ATF/BS units, respectively, were assembled in *BTS1*, *Mcrtl*, and *McrtyB* modules in the Gen 0.1 background (CB1, CB2, and CB3). In strain IMX672.1, weak/medium, strong, and medium ATF/BS units, respectively, were assembled in *BTS1*, *Mcrtl*, and *McrtyB* modules (CC1, CC2, and CC3). We, additionally, tested the effect of strong regulators (CB4 and CC4). Importantly, we only observed yellow colonies, indicating insufficient metabolic flux towards  $\beta$ -carotene formation or activation of alternative pathways. As a further control, modules expressing the three genes from the *TDH3* promoter were integrated in the same loci of Gen 0.1 (**Fig. 5e**) and IMX672.1 (**Fig. 5f**) backgrounds. Almost all colonies growing

on the plate were yellow. The HPLC data (**Fig. 5g**) demonstrate that more  $\beta$ -carotene was produced in the optimized strain than the wild type (CB1,  $0.43 \pm 0.03 \text{ mg g}^{-1} \text{ cdw}$ ; CC1;  $0.41 \pm 0.03 \text{ mg g}^{-1} \text{ cdw}$ ). Moreover, the colonies with ATF/BS control modules produced 1.2- to 1.5-fold more  $\beta$ -carotene than colonies with *TDH3* promoters. We achieved a 1.2-fold improvement of strain CC1 over IME167<sup>32</sup> (see **Supplementary Note 2** and **Supplementary Table 6**).

At Level 2, the pDI-McrtI-BTS1-McrtYB library was transformed into strain Gen 0.1 or IMX672.1 (**Supplementary Note 3**, **Supplementary Fig. 9**). Because colony color may result from the presence of several plasmids within a given yeast cell, the plasmids were recovered from the yeast cells and transformed into *E. coli*. The ATF/BS fragments of five Destination vectors were sequenced. Single plasmids were transformed into strains Gen 0.1 or IMX672.1. The results showed that combining a weak ATF/BS (*BTS1*) with two strong ATF/BSs (*McrtI* and *McrtYB*) results in superior  $\beta$ -carotene accumulation in both strains. Moreover, we observed that expressing all genes from strong ATF/BS units (pDI 5, **Supplementary Fig. 9**) did not result in high  $\beta$ -carotene accumulation. HPLC analysis demonstrated that there is no significant difference in  $\beta$ -carotene production level in the background-optimized strain (G1pDI 1;  $0.81 \pm 0.25 \text{ mg g}^{-1} \text{ cdw}$ ) and the wild type (MXpDI 1;  $0.61 \pm 0.14 \text{ mg g}^{-1} \text{ cdw}$ ).

Next, the Destination vector I library was integrated into the *lys2.a* locus of strains Gen 0.1 or IMX672.1 (**Supplementary Note 4**, **Supplementary Fig. 10**). Weak/medium/strong, weak/strong, and medium ATF/BS regulators, respectively, were employed in the *BTS1*, *McrtI*, and *McrtYB* modules in the Gen 0.1 background, while weak/medium and weak/strong ATF/BS regulators controlled the expression of *BTS1* and *McrtI*, respectively, and all *McrtYB* expressing modules contained strong ATF/BS regulators in the IMX672.1 background. Moreover, 3.3-fold more  $\beta$ -carotene was produced in the background-optimized strain Gen 0.1 (G1intpDI 1) than in the non-optimized IMX672.1 strain (MXintpDI 1).

Overall, the top producer, generated by approach 2, yielded 2.1-fold more  $\beta$ -carotene than IME167<sup>32</sup> (see **Supplementary Note 2** and **Supplementary Table 6**), tested in our experiment. Moreover, our results strongly indicate that a combination of weak, medium, and strong ATF/BSs is required for high-level  $\beta$ -carotene production.



**Fig. 5. Multi-locus integration of  $\beta$ -carotene pathway genes.**

(a) DSB-mediated integration of three-gene  $\beta$ -carotene pathway into three loci. The *McrtI*-, *McrtYB*-, and *BTS1* donors driving  $\beta$ -carotene expression were integrated into the *X-3*, *XI-3*, and *XII-5* loci. Donors contain the library of modules differing in ATF/BS units upstream of  $\beta$ -carotene CDSs and the yeast dominant markers that are flanked by HRs to integrate into the desired loci. Selection on SC-Ura/-Leu/-His media allows screening for successfully integrated cassettes. For simplicity, the IPTG-inducible promoters upstream of the ATF/BSs, terminator fragments, and cleavage sites flanking pathway genes are not included in the figure. (b) Integration of the library of modules containing the  $\beta$ -carotene CDSs under the control of the nine ATF/BS units into strain Gen 0.1 or (c) strain IMX672.1. (d) Identification of ATF/BS units present upstream of each CDS from three different colonies producing deep-orange color in Gen 0.1 (CB1, CB2, and CB3) or IMX672.1 (CC1, CC2, and CC3) backgrounds and modules containing the  $\beta$ -carotene CDSs under the control of strong ATF/BS units in Gen 0.1 (CB4) or IMX672.1 (CC4). The color code is given in **Supplementary Fig. 1**. (e) Modules containing the  $\beta$ -carotene CDSs under the control of the *TDH3* promoter were integrated into strain Gen 0.1 to generate PCB or (f) strain IMX672.1 to generate PCC strains. (g) HPLC analysis of strains CB1, CC1, PCB, and PCC.

**Proof of concept: Building a library for controllable  $\beta$ -ionone production**

We next established a pathway for the biosynthesis of  $\beta$ -ionone in the best  $\beta$ -carotene-producers achieved by the three approaches reported above (**Supplementary Note 5**). RiCCD1 converts  $\beta$ -carotene to  $\beta$ -ionone which leads to yeast cells that are less intensely colored than  $\beta$ -carotene-producing cells. We used approaches 1 (**Supplementary Fig. 11**), 2 (**Supplementary Fig. 12**) and 3 (**Supplementary Fig. 13**), as described in **Supplementary Note 6, 7, and 8**, respectively. We found that approach 3, 1, and 2 resulted in high, medium and low amounts of  $\beta$ -ionone production, respectively. In 86% of the cases (i) medium or high expression of *RiCCD1* was needed to produce a high level of  $\beta$ -ionone in high  $\beta$ -carotene accumulators and (ii) and even more  $\beta$ -ionone accumulated in the IMX672.1 Gen 0.1, demonstrating that superior  $\beta$ -carotene accumulation is not *per se* sufficient for a high-level accumulation of  $\beta$ -ionone in yeast. However, our top  $\beta$ -ionone-producer, strain MXintR1pDI 1 (**Supplementary Fig. 12**), yielded 3.3-fold more  $\beta$ -ionone than RiCCD1<sup>32</sup>, tested in our experiment (see **Supplementary Note 2** and **Supplementary Table 6**).

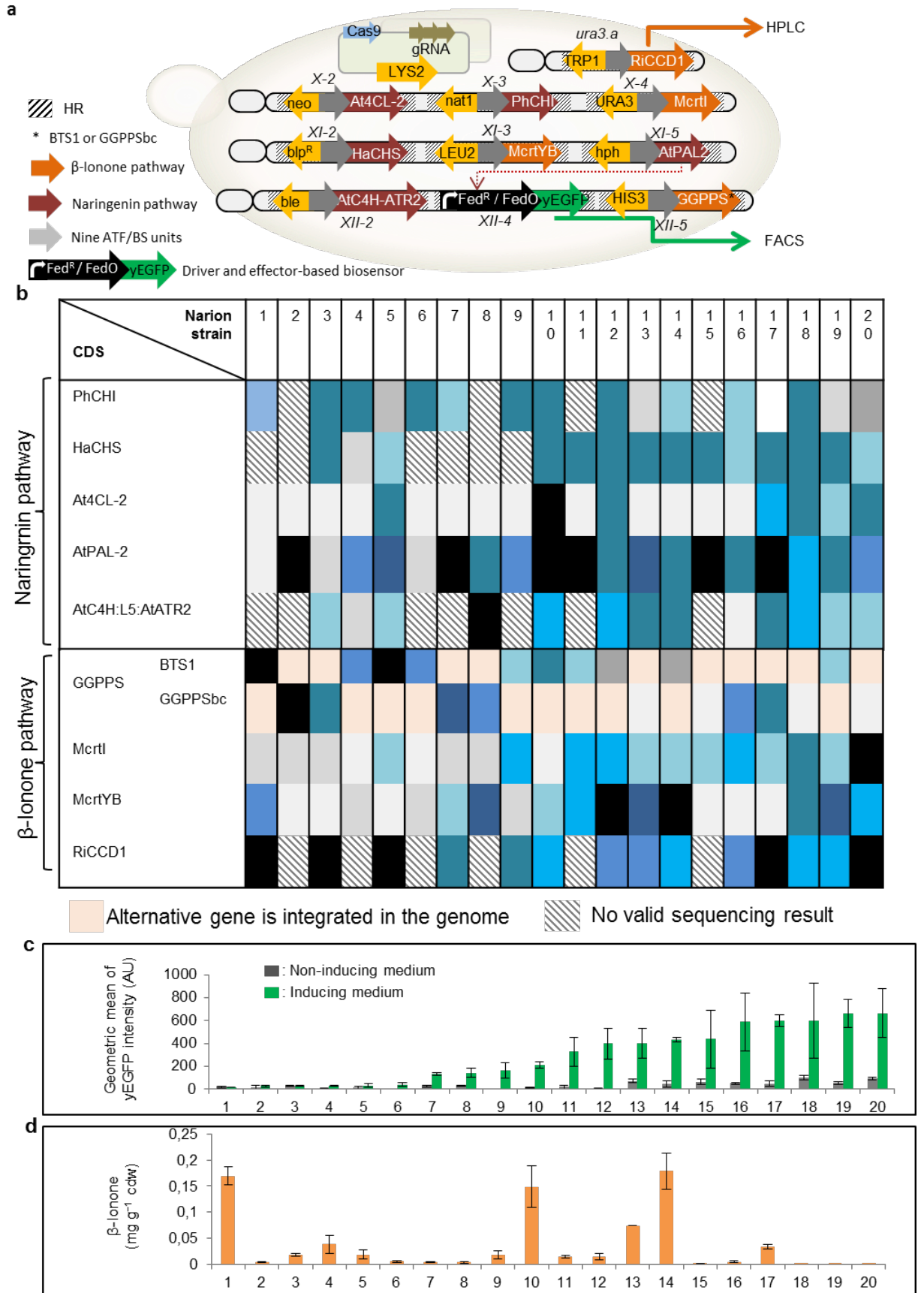
#### **Proof of concept: Co-biosynthesis of $\beta$ -ionone and biosensor-responsive naringenin**

To test the versatility and optimization capacity of COMPASS, we co-engineered the biochemical pathways for the  $\beta$ -ionone (four genes) and naringenin (NG; five genes; *AtC4H:L5:AtATR2*, *PhCHI*, *HaCHS*, *At4CL-2*, and *AtPAL-2*; **Fig. 6**)<sup>24</sup> in *S. cerevisiae*. We detected the accumulation of NG using a recently reported cellular biosensor sensitive to NG<sup>24</sup>. To produce  $\beta$ -ionone, the library of *Mcr1l*, *McrTYB*, *BTS1* (or *GGPPSbc*), and *RiCCD1* CDSs from Level 1 was used (see above and **Online Methods**). For NG biosynthesis, five CDS-terminators of the pathway<sup>24</sup> and promoters of *E. coli* selection markers were assembled in the five Set 2 vectors (**Online Methods** and **Supplementary Protocol**). Subsequently, the five-gene pathway and the nine ATF/BS units were assembled in Destination vector II and Acceptor vectors E - H in a single reaction tube to construct a library of 45 plasmids (see previous chapters). The nine module libraries were integrated into the nine loci of strain MXFde0.2 which harbors the NG biosensor (**Online Methods** and **Supplementary Fig. 14**). The selection for correct integrations was carried out on SC medium containing appropriate selection markers (**Fig. 6a**). The resulting library was called Narion and each colony was expected to contain one out of  $2 \times 9^9$  possible combinations of nine ATF/BS units upstream of the nine CDSs. Twenty colonies (covering 0.0000025% of the theoretical complexity of the library) were selected to identify ATF/BS sequences driving the expression of the CDSs (using primers listed in **Supplementary Table 7**). Of note, all colonies were unique with respect to the combinations of ATF and CDS sequences (**Fig. 6b**). FACS measurements revealed that approximately 30% of the library members show

no or low NG production (Narion 1 – 6; **Fig 6c**). Half of the Narion strains were categorized as midrange producers in which weak and medium ATF/BSs control expression of the NG genes. The better producers (Narion 16 - 20) harbor medium and strong ATF/BS units upstream of NG genes (except Narion 17 and 20 which both contain strong ATF/BS units upstream of *AtPAL-2*). HPLC analyses (**Fig. 6d**) showed that 40% of the library members (Narion 2, 6 - 8, 15 - 16, 19 - 20) produced  $\beta$ -ionone at a level of less than  $0.01 \text{ mg g}^{-1} \text{ cdw}$ , while half of the strains (Narion 3 – 5, 9, 11 – 13, 17 – 18) produced  $0.01 – 0.1 \text{ mg g}^{-1} \text{ cdw}$  of  $\beta$ -ionone. Approximately 15% of the library strains produced  $0.1 – 0.2 \text{ mg g}^{-1} \text{ cdw}$  and harbor a combination of weak, medium and strong ATF/BSs upstream of  $\beta$ -ionone genes. The highest  $\beta$ -ionone yield ( $0.18 \pm 0.017 \text{ mg g}^{-1} \text{ cdw}$ ) was observed in Narion 14. Surprisingly, Narion 19 and 20, top NG accumulators, produced the lowest amount of  $\beta$ -ionone, and Narion 1, producing the lowest amount of NG, was the second-best  $\beta$ -ionone accumulator. Our HPLC data demonstrate that the presence of medium and strong ATF/BS in *Mcr1l* expressing modules result in low-level production of  $\beta$ -ionone. Previously, Ding *et al.* reported that GGPPSbc leads to an improved GGPP supply over BTS1<sup>33</sup>. However, we observed that *BTS1* expressing yeasts produce more  $\beta$ -ionone than *GGPPSbc* expressing cells in 20 characterized Narion strains. Taken together, we identified Narion 14 as the best producer strain as it produces a medium level of NG and the highest level of  $\beta$ -ionone. In Narion 14, the expression of NG and  $\beta$ -ionone pathway genes is controlled by weak/medium and weak/medium/strong ATF/BS regulators, respectively. Overall, through checking 0.0000025% of the theoretical complexity of the library, Narion 14 (**Fig. 6d**) was identified that yielded 4.2-fold more  $\beta$ -ionone than RiCCD1<sup>32</sup>, tested in our experiment (**Supplementary Table 6**)



# COMPASS



**Fig. 6. Illustration of the diversity in  $\beta$ -ionone and NG production from a randomized ATF/BS library.** (a) Schematic overview of the multi-locus integration of  $\beta$ -ionone and NG pathway genes. The *Mcrtl*-, *CrtE*-, *McrtYB*-, *RiCCD1*-, *PhCHI*-, *HaCHS*-, *At4CL-2*-, *AtPAL-2*-, and *AtC4H-ATR2*-CDS donors were integrated into the *X-4*, *XII-5*, *XI-3*, *ura3.a*, *X-3*, *XI-2*, *X-2*, *XI-5*, *XII-2* loci of MXFde0.2, respectively, via CRISPR/cas9-mediated multi-locus integration. Plating cells on SC-Ura/-Leu/-His/-Trp/-Lys media containing G418, hygromycin B, phleomycin, bleomycin, and nourseothricin allows maintaining the Cas9/sgRNA encoding plasmids and screening integrated cassettes. ATFs are expressed from an IPTG-inducible promoter in the presence of inducers (2% (w/v) galactose and 20  $\mu$ M IPTG) and target their BSs upstream of pathway genes.  $\beta$ -Ionone production is quantified by HPLC, while the production of NG is detected using the FdeR biosensor. This results in yEGFP expression detectable by FACS. For simplicity, the IPTG-inducible promoter, terminator, and cleavage sites are not included in the figure. (b) Sequencing results of ATF/BS units present upstream of each CDS. The color code is given in **Supplementary Fig. 1**. (c) Screening of NG production. Twenty colonies were used to monitor yEGFP output in the absence and presence of inducer. Grey, non-induction medium; green, induction medium. Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. AU, arbitrary units. Full data are given in **Supplementary Data 1a**. (d) HPLC analysis for  $\beta$ -ionone production. Data are means  $\pm$  SD ( $n = 3$ ). Full data are given in **Supplementary Data 1b**.

## Discussion

Projects in synthetic biology often require the expression of multiple genes to build biochemical pathways or multiprotein cellular complexes<sup>2</sup>. In a typical scenario, the required expressional activities of the diverse genes are not known. As a consequence, a large number of synthetic constructs where promoters of different strengths drive the expression of the genes are needed. The complexity of such libraries rapidly increases with the number of promoters and genes combined.

Two principal problems need to be solved to allow the establishment and testing such complex construct libraries: (i) A high-throughput method for the reliable assembly of large numbers of diverse constructs is needed<sup>34</sup>. (ii) The product output should be determinable<sup>35</sup>. Colonies are often screened based on the formation of colored products (at least in test cases)<sup>22</sup>, but can also be screened by e.g. HPLC. A more advanced approach employs biosensors able to detect the compounds at the single-cell level<sup>24</sup>.

In the current version of COMPASS, we are able to generate libraries of stable yeast variants with a complexity of theoretically 3,486,784,401 different members through only four cloning reactions followed by the decoupled integration of the constructs into the genome. Of note, to achieve this very large number of constructs we employ only nine of the 106 plant-derived ATF/BS pairs and ten enzyme-encoding open reading frames<sup>9</sup>. The depth at which COMPASS generates diversity is defined by the number of regulators (Y) and open reading frames (N) (size of the library =  $Y^N$ ). To demonstrate a useful application of COMPASS, we generated a library of  $\beta$ -carotene producers; their initial screening proved the power of COMPASS. Our results clearly showed that a combination of different ATF/BSs units, leading to different expression levels of the enzyme-encoding CDSs, was needed for high-level  $\beta$ -carotene production ( $0.81 \pm 0.025$  mg  $\beta$ -carotene  $g^{-1}$  cdw). In our second setup, we co-produced the biosensor-responsive chemical NG and the colorless product  $\beta$ -ionone. Analyzing less than 0.0000025% of the library revealed that approximately 30% of the library members show no or low levels of both chemicals, highlighting the importance of developing combinatorial optimization approaches. Additionally, we found a strain producing  $0.18 \pm 0.017$  mg  $\beta$ -ionone  $g^{-1}$  cdw (4.2-fold more  $\beta$ -ionone than RiCCD1<sup>32</sup>) demonstrating the optimization capacity of COMPASS.

Given the current cyclical and iterative nature of projects targeting the optimization of greater than ten-gene pathways, further improvements may include the utilization of other inducers (e.g. light<sup>36</sup>), adding dynamic regulation to the systems<sup>37</sup>, and adopting COMPASS to a wider range of hosts.

In the present study, we developed the COMPASS toolkit to optimize the production of biochemical compounds. However, besides pathway engineering COMPASS can facilitate many other projects in synthetic biology including e.g. the building of multi-subunit protein complexes, the engineering of sophisticated gene regulatory networks, or the construction of entire synthetic organelles. COMPASS thus has a great potential in many areas of synthetic biology and will help to accelerate the transition from the laboratory to the marketplace.

## Methods

Methods, including statements of data availability and any associated accession codes and references are available in the online version of the paper.

## Acknowledgements

We are thankful to the following colleagues for providing plasmids or yeast strains: Yansheng Zhang (Chinese Academy of Sciences, Beijing, China) for plasmids harboring Mctrl and MctrlYB

coding sequences; Jules Beekwilder (Plant Research International, Wageningen, The Netherlands) for plasmid RiCCD, and the yeast strains IMC167 and RiCCD1; Michael K. Jensen and Tim Snoek (Technical University of Denmark, Lyngby, Denmark) for plasmids pROP280, pROP266, pROP273, pTS-37, pTS-49, and pTAJAK-105; Alex T. Nielsen (Technical University of Denmark, Lyngby, Denmark) for plasmid pTAJAK-92; Verena Siewers (Chalmers University of Technology, Göteborg, Sweden) for yeast strain SCIGS22a; and Ying-jin Yuan (Tianjin University, Tianjin, China) for plasmid pRS425-PTDH3-ts-TPGK-PPGK-GGPPSbc-TCYC1. This work was funded by the Federal Ministry of Education and Research of Germany (BMBF; grant numbers 031A172 and 031B0223, given to Katrin Messerschmidt) and a fellowship of the Potsdam Graduate School, University of Potsdam, given to GN. All experiments were done in the Cell2Fab lab, led by KM. HPLC analyses were ordered by KM and were performed by AppliChrom® (Oranienburg, Germany).

### Author contributions

GN conceived and designed the COMPASS strategy, performed the experiments and analyzed the data. JB and LR helped to construct the naringenin plasmids and strains. BMR supervised the research conducted by GN and provided expert knowledge. GN and BMR wrote the manuscript. All authors take full responsibility for the content of the paper.

### Competing financial interests

The authors declare no competing financial interests.

### References

1. Guo, Y. *et al.* YeastFab: the design and construction of standard biological parts for metabolic engineering in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **43**, e88 (2015).
2. Khalil, A.S. *et al.* A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647-658 (2012).
3. Ling Yuan, S.E.P. Plant transcription factors: methods and protocols. ( Humana Press, New York; 2011).
4. Mclsaac, R.S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res* **41**, e57 (2013).
5. Brophy, J.A. & Voigt, C.A. Principles of genetic circuit design. *Nat Methods* **11**, 508-520 (2014).
6. Reider Apel, A. *et al.* A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **45**, 496-508 (2017).

7. Naseri, G. *et al.* Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth Biol*, **6**,1742-1756 (2017).
8. Pflieger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* **24**, 1027-1032 (2006).
9. Dueber, J.E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* **27**, 753-759 (2009).
10. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946-950 (2009).
11. Anthony, J.R. *et al.* Optimization of the mevalonate-based isoprenoid biosynthetic pathway in *Escherichia coli* for production of the anti-malarial drug precursor amorpho-4,11-diene. *Metab Eng* **11**, 13-19 (2009).
12. Bond-Watts, B.B., Bellerose, R.J. & Chang, M.C. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nat Chem Biol* **7**, 222-227 (2011).
13. Shen, C.R. *et al.* Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Appl Environ Microbiol* **77**, 2905-2915 (2011).
14. Lian, J., Hamedirad, M., Hu, S. & Zhao, H. Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. *Nat Commun* **8**, 1688 (2017).
15. Du, J., Yuan, Y., Si, T., Lian, J. & Zhao, H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res* **40**, e142 (2012).
16. Borodina, I. & Nielsen, J. Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnol J* **9**, 609-620 (2014).
17. Holtz, W.J. & Keasling, J.D. Engineering static and dynamic control of synthetic pathways. *Cell* **140**, 19-23 (2010).
18. Nole-Wilson, S. & Krizek, B.A. DNA binding properties of the Arabidopsis floral development protein Aintegumenta. *Nucleic Acids Research* **28**, 4076-4082 (2000).
19. Ellis, T., Wang, X. & Collins, J.J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol* **27**, 465-471 (2009).
20. Baek, C.H., Chesnut, J. & Katzen, F. Positive selection improves the efficiency of DNA assembly. *Anal Biochem* **476**, 1-4 (2015).
21. Futcher, B. & Carbon, J. Toxic effects of excess cloned centromeres. *Molecular and cellular biology* **6**, 2213-2222 (1986).
22. Mitchell, L.A. *et al.* Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Res* **43**, 6620-6630 (2015).
23. Lopez, J. *et al.* Production of beta-ionone by combined expression of carotenogenic and plant CCD1 genes in *Saccharomyces cerevisiae*. *Microb Cell Fact* **14**, 84 (2015).
24. Skjoedt, M.L. *et al.* Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nat Chem Biol* **12**, 951-958 (2016).

25. Bao, Z. *et al.* Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS Synth Biol* **4**, 585-594 (2015).
26. Liang, J., Liu, Z., Low, X.Z., Ang, E.L. & Zhao, H. Twin-primer non-enzymatic DNA assembly: an efficient and accurate multi-part DNA assembly method. *Nucleic Acids Res* (2017).
27. Kouprina, N. & Larionov, V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* **125**, 621-632 (2016).
28. Zwolshen, J.H.a.B., J. K. Genetic and biochemical properties of thialysine-resistant mutants of *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* **122**, 281– 287 (1981).
29. Dorfman, B.Z. The isolation of adenylosuccinate synthetase mutants in yeast by selection for constitutive behavior in pigmented strains *Genetics* **61**, 377– 389 ( 1969).
30. Li, Q., Sun, Z., Li, J. & Zhang, Y. Enhancing beta-carotene production in *Saccharomyces cerevisiae* by metabolic engineering. *FEMS Microbiol Lett* **345**, 94-101 (2013).
31. Lee, M.E., DeLoache, W.C., Cervantes, B. & Dueber, J.E. A Highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth Biol* **4**, 975-986 (2015).
32. Beekwilder, J. *et al.* Polycistronic expression of a beta-carotene biosynthetic pathway in *Saccharomyces cerevisiae* coupled to beta-ionone production. *J Biotechnol* **192 Pt B**, 383-392 (2014).
33. Ding, M.Z. *et al.* Biosynthesis of Taxadiene in *Saccharomyces cerevisiae* : selection of geranylgeranyl diphosphate synthase directed by a computer-aided docking strategy. *PLoS One* **9**, e109348 (2014).
34. Jeschek, M., Gerngross, D. & Panke, S. Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nat Commun* **7**, 11163 (2016).
35. Zhang, J., Jensen, M.K. & Keasling, J.D. Development of biosensors and their application in metabolic engineering. *Curr Opin Chem Biol* **28**, 1-8 (2015).
36. Gao, Y. *et al.* Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat Methods* **13**, 1043-1049 (2016).
37. Siciliano, V. *et al.* Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line. *PLoS Comput Biol* **7**, e1002074 (2011).

## Online Methods

### General

The list of *Saccharomyces cerevisiae* strains used in this study and their genotypes are given in **Supplementary Table 5**. Plasmids were constructed by NEBuilder HiFi DNA assembly (New England Biolabs, Frankfurt am Main, Germany) and SLiCE cloning<sup>1, 2</sup>. Plasmids and primer sequences are given in **Supplementary Table 8** and **9**. PCR amplifications were done using Phusion (Thermo Fisher Scientific), Q5 DNA Polymerases (New England Biolabs) or PrimeSTAR GXL DNA Polymerase (Takara Bio, Saint-Germain-en-Laye, France) according to the manufacturer's recommendations. Amplified DNA parts were gel-purified prior to further use. All primers and oligonucleotides were ordered from Eurofins Genomics (Ebersberg, Germany). All constructs were confirmed by sequencing (LGC Genomics, Berlin, Germany).

### Bacterial and yeast strains

Plasmids were transformed into *Escherichia coli* NEB 5 $\alpha$  or NEB 10 $\beta$  cells (New England Biolabs), or into ElectroSHOX Competent Cells (Biolone, Luckenwalde, Germany). Strains were grown in Luria-Bertani medium with appropriate selection marker at 28°C (triclosan, 14.5  $\mu$ g/ml) or 37°C (spectinomycin, 50  $\mu$ g/ml; ampicillin, 50  $\mu$ g/ml; chloramphenicol, 25  $\mu$ g/ml; gentamicin, 50  $\mu$ g/ml).

*S. cerevisiae* strains YPH500 (ATCC: 76626), IMX672 (Euroscarf, #Y40595), and SCIGS22a<sup>3</sup> were used. Information regarding yeast strains constructed in this work is presented in **Supplementary Table 5**. Generation of competent yeast cells and genetic transformation of plasmids or linearized DNA fragments were done using either the LiAc/SS carrier DNA/PEG method<sup>4</sup> or The Frozen-EZ Yeast Transformation II Kit (Zymo Research, Freiburg, Germany). All strains were grown at 30°C in yeast extract peptone dextrose adenine (YPDA)-rich medium or in appropriate synthetic complete (SC) media lacking one or more amino acids to allow selection for transformed cells. Dominant selection markers were used in YPDA medium in the indicated final concentrations: G418 (200  $\mu$ g/ml), hygromycin B (200  $\mu$ g/ml), phleomycin (20  $\mu$ g/ml), bleomycin (100  $\mu$ g/ml), and nourseothricin (100  $\mu$ g/ml). Verification of transformation was done using colony PCR followed by sequencing. Either Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research) or the method described by Noskov *et al.*<sup>5</sup> were used to recover plasmid DNA from positive yeast clones.

**Construction of plasmids for strain generation**

**pCOM001:** To construct pCOM001, plasmids pCOMA and pCOMB were ordered from MWG to PCR-amplify fragments A (primers COMA\_for/COMA\_rev, on pCOMA) and B (primers COMB\_for/COMB\_rev, on pCOMB). Next, they were individually assembled into *Bsal*-digested pCRCT plasmid (Addgene #60621). Resulting plasmids were named pCOMC and pCOMD. Subsequently, PCR- amplified fragment (primers COMD\_for/COMD\_rev, on pCOMD) was assembled into *Bsal*-digested pCOMC. Resulting plasmid was named pCOM001.

**pCOM002:** To construct pCOM002, PCR- amplified fragment neo gene (primers KAN\_for/KAN\_rev, on pTAJAK92<sup>6</sup>), fragment containing *CYC1* terminator - structural RNA - *STU4* terminator (tCYC1\_for/tCYC1\_rev, on BY4741 genomic DNA), and *SNR52* fragment (primers SNR\_for/SNR\_rev, on pCRCT) were assembled into *Bsal*-digested pCOM001. Resulting plasmid was named pCOM002.

**pCOM003:** The *CEN/ARS* origin of replication (CEN\_for/CEN\_rev, on pGN006<sup>7</sup>) was cloned into *PmeI*-digested pGN003B<sup>7</sup> to construct pGN003BM. Next, plasmid pCOM003 was constructed by Gibson assembly, amplifying the Cre-EBD transcription unit from pDL12<sup>8</sup> and its insertion into the *EcoRI/SacII*-digested pGN003BM<sup>7</sup>.

**pCOM004:** The *LYS2* encoding fragment (LYSA\_for/LYSA\_rev, on pYC6Lys-TRP1URA3, Addgene # 11010) and iCas9 encoding fragment (CASA\_for/CASA2\_rev, on pCRCT) were cloned into *NcoI/NotI*-digested pTAJAK-92<sup>6</sup>.

**pCOM005:** The *LYS2* encoding fragment (LYSA\_for/LYSA\_rev, on pYC6Lys-TRP1URA3, Addgene #11010) and iCas9 encoding fragment (CASA\_for/CASA\_rev, on pCRCT) were cloned into *NcoI/SphI*-digested pCfB3052 (Addgene #73294).

**pCOM006:** The *LYS2* encoding fragment (LYSB\_for/LYSA\_rev, on *pYC6Lys-TRP1URA3*, Addgene #11010), iCas9 encoding fragment (CASA\_for/CASB\_rev, on pCRCT), and *Hph*<sup>R</sup> encoding fragment (HYG\_for/HYG\_rev, on Acceptor vector H, see COMPASS vector section) were cloned into *SfoI*-digested pTAJAK-105<sup>6</sup>.

**pCOM007:** The *LYS2* encoding fragment (LYSA\_for/LYSA\_rev, on pYC6Lys-TRP1URA3, Addgene #11010) and iCas9 encoding fragment (CASA\_for/CASA\_rev, on pCRCT) were cloned into *NcoI/SphI*-digested pCfB3051 (Addgene #73293).

**pCOM008:** The *LYS2* encoding fragment (LYSA\_for/LYSA\_rev, on pYC6Lys-TRP1URA3, Addgene #11010) and *iCas9* encoding fragment (CASA\_for/CASA\_rev, on pCRCT) were cloned into *NcoI/SphI*-digested pCfB3053 (Addgene #73295).

**Construction of yeast strains**



Strain Gen 0.1: The *S. cerevisiae* strain SCIGS22a<sup>3,9</sup> was used in this work to generate strain Gen 0.1. Strain SCIGS22a has a CEN.PK background with additional modifications in the genome for the overaccumulation of FPP<sup>3,9</sup> and is auxotrophic for *URA3*. COMPASS employs a positive selection scheme that involves five auxotrophic marker genes. The marker's coding sequences are provided within the vector backbones, while their corresponding promoters are part of the assembly fragments used in the TAR reaction. The promoter drives expression of the corresponding auxotrophic marker only in successfully assembled constructs. Therefore, SCIGS22a needed to be auxotrophic for *HIS3*, *LEU2*, *TRP1*, and *LYS2*, in addition to *URA3*. Hence, we used the Homology-Integrated CRISPR-Cas (HI-CRISPR) system and its design principles for one-step multiple auxotrophic gene disruption<sup>10</sup>. HI-CRISPR uses plasmid pCRCT (Addgene, #60621), a high-copy plasmid harboring iCas9, a variant of wild-type Cas9 that increases the gene disruption efficiency, trans-encoded RNA (tracrRNA), and a homology-integrated crRNA cassette. It relies on the insertion of a 100-bp dsDNA mutagenizing homologous recombination donor between two direct repeats in the case of each target gene. Multiple donors and corresponding guide sequences can be introduced in pCRCT. For each gene disruption, a gRNA targeting 20-bp unique sequence was selected via BLAST searches against the *S. cerevisiae* S288c genome (NCBI Taxonomy ID: 559292) to minimize off-target effects. In the next step, plasmid pCOM001 for quadruple auxotrophic gene disruption (see above and **Supplementary Fig. 14a**) and transformed into *S. cerevisiae* SCIGS22a cells. Hence, *leu2.a* (519-bp downstream of the *LEU2* start codon), *his3.a* (265-bp downstream of the *HIS3* start codon), *lys2.a* (799-bp downstream of the *LYS2* start codon), and *trp1.a* (245-bp downstream of the *TRP1* start codon) were targeted to achieve frame-shift mutations. The transformed cells were inoculated in liquid SC-Ura culture overnight. After 4 d, 200  $\mu$ l of a 10<sup>4</sup>-fold diluted cell culture were plated on SC-Ura plates. After 2 d, a total of 50 colonies were randomly selected and each single colony was streaked out onto four different selective plates (*i.e.*, SC-Leu, SC-His, SC-Lys, SC-Trp). After two more days, cells of a colony that did not grow on either of the four selective plates were streaked out on non-selective YPDA agar medium to eliminate the pCOM001 plasmid. After four rounds of re-streaking single colonies onto new YPDA plates, we recovered a colony that was also not able to grow on SC-Ura medium. The new strain was named Gen 0.1.

Strain IMX672.1: To achieve a *lys2.a* frame-shift mutation in strain IMX672 (derived from strain CEN.PK), plasmid pCOM001 was transformed into the strain and selection was done as described for Gen 0.1 above. Thirty colonies were randomly streaked out on SC-Lys and YPDA plates. The colony that did not grow on SC-Lys plate carries the *lys2.a* frame shift mutation. To remove pCOM001, the corresponding colony was re-streaked several rounds on YPDA and SC-

Ura plates, which yielded a colony that was not able to grow on SC-Ura. The corresponding strain was named IMX672.1.

Strain MXFde 0.2: The *S. cerevisiae* strain TSINO 93 (FdeR-based reporter strain)<sup>11</sup> has a CEN.PK background with *Pro<sub>TDH3</sub>-FdeR-URA3* and *Pro<sub>CYC1-FdeO</sub>-GFP-LoxP-HphMX-LoxP* genes, and is auxotrophic for *LEU2* and *TRP1*. COMPASS employs a positive selection scheme that involves auxotrophic and dominant marker genes. Therefore, TSINO 93 needed to be auxotrophic for *HIS3*, *URA3*, and *LYS2*, in addition to *LEU2* and *TRP1*, and the *HphMX* dominant marker needed to be deleted from the strain. We employed the HI-CRISPR method for one-step multi-gene disruption<sup>10</sup>. Therefore, fragment encoding gRNA to target 133-bp downstream of the *URA3* start codon (*ura3.a* mutation) were introduced in pCOM001 plasmid (**Supplementary Fig. 14b**). The resulting plasmid was called pCOM002 (see above) and co-transformed with the *ura3.a* donor (annealed single-stranded oligonucleotides URA3.A\_for/URA3.A) into TSINO93 strain to achieve frame-shift mutations. Two-hundred microliters of a 10<sup>4</sup>-fold diluted cell culture were plated on YPDA plate with G418. After 2 d, a total of 50 colonies were randomly selected and each single colony was streaked out onto three different selective plates (*i.e.*, SC-His/-Lys/-Ura). After several rounds of re-streaking single colonies onto new YPDA plates, we recovered a colony that was also not able to grow on YPDA medium with nourseothricin. After two more days, cells of a colony that did not grow on either of the three selective plates were streaked out on non-selective YPDA agar medium to eliminate the pCOM002 plasmid. The new strain was named MXFde 0.1. To generate strain MXFde 0.2, we employed  $\beta$ -estradiol (EST)-induced Cre recombinase<sup>12</sup> to remove the *HphMX* CDS flanked by *loxPsym* sites<sup>11</sup>. Plasmid pCOM003 was constructed as described and transformed into MXFde 0.1 cells. An EST-Cre cell culture was plated on SC-Trp medium. After 3 d, single colonies were inoculated in liquid SC-Trp medium and grown in darkness for 6 h at 30°C and 230 rpm, then induced with 2  $\mu$ M  $\beta$ -estradiol (Sigma-Aldrich, Munich, Germany) and grown for another 24 h. Two-hundred microliters of a 10<sup>4</sup>-fold diluted cell culture were plated on YPDA plates containing 200  $\mu$ g/ml hygromycin B. After several rounds of re-streaking single colonies onto new YPDA plates, we recovered a colony that was also not able to grow on YPDA medium with hygromycin B. The positive colony was checked by PCR for deletion of the *HphMX* CDS. After two more days, cells of a colony that did not grow on selective plates were streaked out on non-selective YPDA agar medium to eliminate the pCOM003 plasmid. The new strain was named MXFde 0.2.

### Construction of COMPASS plasmids

All cloning steps are performed using overlap-based methods. To this end, regions homologous to neighboring fragments are included in the primers used to PCR amplify the DNA fragments to be joined.

**Entry vector X:** pGN003B<sup>7</sup> was digested with *PmeI*, treated with T4 DNA polymerase to remove 3' overhang end and religated to construct pCOMPASS01. Next, pLOA\_0\_1<sup>13</sup> was cut with *AscI/FseI* and used in an assembly reaction using PCR amplified *Pro<sub>TEF1</sub> - LacI - Ter<sub>ADH1</sub>* (primers LAC\_for/LAC\_rev, on pCOMPASS01) to generate pCOMPASS02. *NotI/PacI*-digested pCOMPASS02 was used in an assembly reaction using PCR amplified *Pro<sub>TRP1</sub>* (primers PTRP\_for/PTRP\_rev, on BY4741 genomic DNA). Thereby, *Bam*HI and *Sal*I sites were introduced downstream of *Pro<sub>TRP1</sub>*, while *PacI* site was introduced upstream of *Pro<sub>TRP1</sub>*. The resulting plasmid was called pCOMPASS03. The pCOMPASS03 plasmid was digested with *SbfI*, treated with T4 DNA polymerase to remove 3' overhang end and religated. The resulting plasmid was called pCOMPASS04. *Bam*HI/*Sal*I-digested pCOMPASS04 used in an assembly reaction with annealed single-stranded oligonucleotides CYCM\_for/CYCM\_rev introducing MCS and X0, the last 30 bp of *Pro<sub>CYC1mini</sub>*, between *Pro<sub>TRP1</sub>* and *E. coli ori*. The resulting plasmid was called Entry vector X.

**Acceptor vectors A - D:** To construct the first set (Set 1) of Acceptor vectors containing auxotrophic selection markers, *Bam*HI/*NotI*-digested pCOMPASS04 was used in an assembly reaction using PCR-amplified DNA parts as follows:

- i. pLOA\_0\_1<sup>13</sup> was cut with *NotI*. A two-way Gibson cloning of *Pro<sub>KanaR</sub>* (primers PKANA\_for/PKANA\_rev, on pCR4-topo, Invitrogen, Karlsruhe, Germany), *Amp<sup>R</sup>* (primers AMP\_for/AMP\_rev, on pGN005B<sup>7</sup>) and *Ter<sub>KanaR</sub>* (primers TKANA\_for/TKANA\_rev, on pCR4-topo, Invitrogen) were done. The resulting plasmid was called pCOMPASS06. *HIS3* and *Ter<sub>HIS3</sub>* (primers HISTERA\_for/HISTERA\_rev, on pGN005B<sup>7</sup>), *Pro<sub>HIS3</sub>* (primer PHIS\_for/PHIS\_rev, on pGN005B<sup>7</sup>), *Amp<sup>R</sup>* and *Ter<sub>KanaR</sub>* (primers AMPTER\_for/AMPTER\_rev, on pCOMPASS06) to result in Acceptor vector A.
- ii. *LEU2* and *Ter<sub>LEU2</sub>* (primers LEUTER\_for/LEUTER\_rev, on pGAD424, TAKARA Bio, GenBank #U07647), *Pro<sub>LEU2</sub>* (primer PLEU\_for/PLEU\_rev, on pGAD424, TAKARA Bio, GenBank #U07647), *Cm<sup>R</sup>* and *Ter<sub>CmR</sub>* (primers CMRTER\_for/CMRTER\_rev, on pLD\_3\_4<sup>13</sup>) to result in Acceptor vector B.
- iii. *TRP1* and *Ter<sub>TRP1</sub>* (primer TRPTEC\_for/TRPTEC\_rev, on pGN003B<sup>7</sup>), *Pro<sub>TRP1</sub>* (primers PTRP\_for/PTRP\_rev, on pGN003B<sup>7</sup>), *TCS<sup>R</sup>* and *Ter<sub>TCSR</sub>* (primers TCSRTER\_for/TCSRTER\_rev, on pF2, Addgene #42520) to result in Acceptor vector C.

- iv. *LYS2* and *Ter<sub>LYS2</sub>* (primers LYSTERA\_for/LYSTERA\_rev, on pYC6Lys-TRP1URA3, Addgene #11010), *Pro<sub>LYS1</sub>* (primers PLYS\_for/PLYS\_ter, on pYC6Lys-TRP1URA3, Addgene #11010), *Gen<sup>R</sup>* and *Ter<sub>GenR</sub>* (primers GENTER\_for/GENTER\_rev, on pDEST321) to result in Acceptor vector D.

**Destination vector I:** pCOMPASS03 was digested with *PacI* and *Clal* to remove *Pro<sub>URA3</sub>* - *URA3* - *Ter<sub>URA3</sub>* from downstream (right) of *E. coli<sub>ori</sub>*. The ~5-kb vector was then used in assembly reaction using PCR-amplified fragment (primers MURA\_for/MURA\_rev, on pCOMPASS03). The resulting plasmid was called pCOMPASS07. To construct pCOMPASS08, *NotI/PacI*-digested pCOMPASS07 was used in an assembly reaction using PCR-amplified *Pro<sub>URA3</sub>* (primers URATERI\_for/URATERI\_rev, on pCOMPASS04) and PCR-amplified *URA3* - *Ter<sub>URA3</sub>* (primer PURAI\_for/PURAI\_rev, on pCOMPASS04). Therefore, a *NotI* site was introduced between *Pro<sub>URA3</sub>* and the *URA3* CDS, while *BamHI* and *PacI* were inserted downstream of the *X0* site. To construct pCOMPASS09, *BamHI/PacI*- digested pCOMPASS08 was used in an assembly reaction using *Spect<sup>R</sup>* - *Ter<sub>SpectR</sub>* (primers SPECTERI\_for/SPECTTERI\_rev, on pCR8/GW/TOPO, TOPO Cloning Kit, TAKARA). Thereby, seven nucleotides were removed from upstream of *Spect<sup>R</sup>* and *PacI* site were inserted downstream of *Ter<sub>SpectR</sub>*. Moreover, *BamHI/SalI*- digested pCOMPASS04 was used in a two-way assembly reaction using PCR-amplified DNA parts as follows:

- i. *LYS2* and *Ter<sub>LYS2</sub>* (primers LYSTERI\_for/LYSTERI\_rev, on pYC6Lys-TRP1URA3, Addgene #11010), and annealed single-stranded oligonucleotides LYSX0\_for/LYSX0\_rev to result in pCOMPASS10. Thereby, the *X1* sequence, *AscI* site, and *Y4* were introduced between *LYS2* CDS and *PacI* site.
- ii. *TRP1* and *Ter<sub>TRP1</sub>* (primer TRPTERI\_for/TRPTERI\_rev, on pGN003B<sup>7</sup>), and annealed single-stranded oligonucleotides TRPX0\_for/TRPX0\_rev to result in COMPASS11. Thereby, *X1* sequence, *SfiI* and *Pi-PspI* sites, and *Y3* were introduced between *LYS2* CDS and *PacI* site.
- iii. *LEU2* and *Ter<sub>LEU2</sub>* (primers LEUTERI\_for/LEUTERI\_rev, on pGAD424, TAKARA Bio, GenBank #U07647), and annealed single-stranded oligonucleotides LEUX0\_for/LEUX0\_rev to result in COMPASS12. Thereby, *X1* sequence, *FseI* and *I-SceI* sites, and *Y2* were introduced between *LEU2* CDS and *PacI* site.
- iv. *HIS3* and *Ter<sub>HIS3</sub>* (primers HISTERI\_for/HISTERI\_rev, on pGN005B<sup>7</sup>) and annealed single-stranded oligonucleotides HISX0\_for/HISX0\_rev introducing homology region to result in COMPASS13. Thereby, *X1* sequence, *SbfI* and *I-CeuI* sites, and *Y1* were introduced between the *HIS3* CDS and the *PacI* site.

Next, to construct pCOMPASS14, *PacI*- digested pCOMPASS09 was used in an assembly reaction with PCR-amplified fragment *Ter<sub>HIS3</sub> - HIS3 - X1 - I-CeuI - Y1* (primers HIS\_for/HIS\_rev, on COMPASS13). To construct pCOMPASS15, *FseI/PacI*-digested *pCOMPASS14* was used in an assembly reaction with PCR-amplified fragment *Ter<sub>LEU2</sub> - LEU2 - X1 - FseI - I-SceI - Y2* (primers LEU\_for/LEU\_rev, on pCOMPASS12). To construct pCOMPASS16, *AscI/PacI*-digested COMPASS15 was used in an assembly reaction using PCR-amplified fragment *Ter<sub>TRP1</sub> - TRP1 - X1- SfiI - PI-PspI -Y3* (primers TRP\_for/TRP\_rev, on pCOMPASS11). To construct Destination vector I, *PacI*-digested pCOMPASS15 was used in an assembly reaction with PCR-amplified fragment *Ter<sub>TRP1</sub> - TRP1 - X1 - SfiI - PI-PspI - Y3 - Ter<sub>LYS2</sub> - LYS2 - X1 - AscI - Y4* (primers TRPI\_for/LYSI\_rev, on pCOMPASS16). To construct Destination vector I.1, annealed single-stranded oligonucleotide LYP\_for/LYP\_rev was digested with *PacI* and was used to ligate in *PacI*-digested Destination vector I. The resulting plasmid was called Destination vector I.1. Through this, 45-bp *LHR - PmeI - AatII*- 45bp *R<sub>HR</sub>* were introduced into the plasmid to integrated Destination vector I.1 digested with either *PmeI* or *AatII*-digested in *LYP1.x*<sup>10</sup> site of yeast genome.

#### Acceptor vectors E - H:

To construct the second set (Set 2) of Acceptor vectors, containing dominant resistance markers, *BamHI/NotI*- digested pCOMPASS04 was used in a three-way assembly reaction using PCR-amplified DNA parts as follows:

- i. *Nat1* and *Ter<sub>FBA1</sub>* (primers NATRER\_for/NATRER\_rev, on pAG36, Addgene #35126), *Pro<sub>FBA1</sub>* (primers PFBA\_for/PFAB\_rev, on pAG36, Addgene #35126) and *Amp<sup>R</sup>* and *Ter<sub>AmpR</sub>* (primers AMPTERE\_for/AMPTERE\_rev, on pCOMPASS06) to result in pCOMPASS17. Next, pCOMPASS17 was digested with *PciI* and *NdeI* to delete 264-bp from the *URA3* encoding sequence. The remaining ~6-kb fragment was used in an assembly reaction using annealed single-stranded oligonucleotides DURA\_for/DURA\_rev. The resulting plasmid was named Acceptor vector E.
- ii. *Ble* and *Ter<sub>ble</sub>* (primers BLERTER\_for/BLERTER\_rev, on pCEV-G1-Ph, Addgene #46814), *Pro<sub>PGK1</sub>* (primer PGKF\_for/PGKF\_rev, on pCEV-G1-Ph, Addgene #46814), and *Cm<sup>R</sup>* and *Ter<sub>CmR</sub>* (CMRTERF\_for/CMRTERF\_rev, on pLD\_3\_4<sup>13</sup>) to result pCOMPASS018. Next, pCOMPASS018 was digested with *PciI* and *NdeI* to delete 273-bp from the *URA3* encoding sequence. The remaining ~6-kb fragment was used in an assembly reaction using annealed single-stranded oligonucleotides DURA\_for/DURA\_rev. The resulting plasmid was named Acceptor vector F.

- iii. *Neo* and *Ter<sub>neo</sub>* (primers KANARTER\_for/KANARTER\_rev, on pCEV-G2-Km, Addgene #46815), *Pro<sub>TDH3</sub>* (primers TDHG\_for/TDHG\_rev, on pCEV-G2-Km, Addgene #46815), *TCS<sup>R</sup>* and *Ter<sub>TCSR</sub>* (primers, on TCSRGTER\_for/TCSRGTER\_rev, on pF2, Addgene #42520) to result in pCOMPASS018. Next, pCOMPASS019 was digested with *PciI* and *NdeI* to delete 273-bp from the *URA3* encoding sequence. The remaining ~6-kb fragment was used in assembly reaction using annealed single-stranded oligonucleotides DURA\_for/DURA\_rev. The resulting plasmid was named Acceptor vector G.
- iv. *Hph* and *Ter<sub>hph</sub>* (primers HPHRTER\_for/HPHRTER\_rev, on pHIS3p:mRuby2-Tub1+3'UTR::HPH, Addgene #50633), *Pro<sub>TEF1</sub>* (primers PTEFH\_rev/PTEFH\_rev, pHIS3p:mRuby2-Tub1+3'UTR::HPH, Addgene #50633), *Gen<sup>R</sup>* and *Ter<sub>GenR</sub>* (primers GENTERH\_for/GENTERH\_rev, TAKARA Bio) to result in pCOMPASS20. Next, pCOMPASS020 was digested with *AcI* and *Clal* to delete 974-bp including *Pro<sub>URA3</sub>* and the first 749-bp of the *URA3* CDS. The remaining ~5.5-kb fragment was used in an assembly reaction using annealed single-stranded oligonucleotides DDURA\_for/DDURA\_rev. The resulting plasmid was named Acceptor vector H.

#### Destination vector II:

Destination vector II was constructed in the following way.

- i. *NotI/I-CeuI*-digested pCOMPASS13 was used in an assembly reaction with PCR-amplified *nat1* and *Ter<sub>nat1</sub>* (primers NATRERII\_for/NATRERII\_rev, on pAG36, Addgene #35126) to result in pCOMPASS21. Thereby, *X1* sequence, *SbfI* and *I-CeuI* sites, and the *Y1* sequence were introduced between the *nat1* CDS and the *PacI* site.
- ii. *FseI/NotI*-digested pCOMPASS12 was used in an assembly reaction with PCR-amplified *ble* and *Ter<sub>ble</sub>* (primers BLERTERII\_for/BLERTERII\_rev, on *pCEV-G1-Ph*, Addgene #46814) to result in pCOMPASS22. Thereby, *X1* sequence, *FseI* and *I-SceI* sites, and *Y2* were introduced between the *ble* CDS and the *PacI* site.
- iii. *PI-PspI/NotI*-digested pCOMPASS11 was used in an assembly reaction with PCR-amplified *neo* and *Ter<sub>neo</sub>* (primers KANARTERII\_for/KANARTERII\_rev, on pCEV-G2-Km, Addgene #46815) to result in pCOMPASS23. Thereby, *X1* sequence, *SfiI* and *PI-PspI* sites, and *Y3* were introduced between the *neo* CDS and the *PacI* site.
- iv. *AscI/NotI*-digested pCOMPASS10 was used in an assembly reaction with PCR-amplified *hph* and *Ter<sub>hph</sub>* (primers HPHRTERII\_for/HPHRTERII\_rev, on pHIS3p:mRuby2-Tub1+3'UTR::HPH, Addgene #50633) to result in pCOMPASS24. Thereby, *X1*, *AscI* sites, and *Y4* were introduced between the *Hph* CDS and the *PacI* site.

Next, pCOMPASS07 was digested using *NotI* and *PacI* to remove *Pro<sub>Trp1</sub>*. The remaining ~5.2-kb fragment was used in an assembly reaction using with PCR-amplified *Spect<sup>R</sup>* and *Ter<sub>Spect<sup>R</sup></sub>* (primers SPECTII\_for/SPECTII\_rev, on pCOMPASS09) to result in pCOMPASS25. *PacI*-digested pCOMPASS05 was used in an assembly reaction with PCR-amplified fragment *nat1 – Ter<sub>nat1</sub> – X1 - SbfI - I-CeuI – Y1* (primers NATII\_for/NATII\_rev, on COMPASS21). The resulting plasmid was called pCOMPASS26. To construct pCOMPASS27, *FseI/PacI*-digested pCOMPASS26 was used in an assembly reaction with PCR-amplified fragment *Ble – Ter<sub>ble</sub> – X1 - FseI - I-SceI– Y2* (primers BLEII\_for/BLEII\_rev, on COMPASS22). To construct pCOMPASS28, *FseI/PacI*-digested pCOMPASS23 was used in an assembly reaction with PCR-amplified fragment *hph – Ter<sub>hph</sub> – X1- AscI – Y3* (primers HPHRII\_for/HPHRII\_rev, pCOMPASS27). To construct pCOMPASS29, *PacI*-digested pCOMPASS27 was used in an assembly reaction with PCR-amplified fragment *Hph – Ter<sub>hph</sub> – AscI – Y3 - neo – Ter<sub>neo</sub> – X1 - SfiI - PI-PspI – Y4* (HPHRII\_for/HKANARII\_rev, on pCOMPASS28). *NotI*-digested pCOMPASS29 was used in an assembly reaction with annealed single-stranded oligonucleotides RE\_for/RE\_rev. Through this, *Bam*HI and *Xho*I site were introduced between the *X0* sequence and the *Spect<sup>R</sup>* CDS. The resulting plasmid was called pCOMPASS30. *NotI*-digested pCOMPASS30 was used in an assembly reaction with PCR-amplified *Pro<sub>TEF1</sub> – blp<sup>R</sup> – Ter<sub>TEF1</sub>* (BLPR\_for/BLPR\_rev, on pAG31, Addgene #35124). Through this, a *NotI* site was introduced between the *blp<sup>R</sup>* encoding fragment and the *X0* sequence. The resulting plasmid was called pCOMPASS31. Finally, oligonucleotides ADE2A\_for and ADE2A\_rev were annealed and the resulting double-strand oligonucleotide was digested with *PacI* and ligated into *PacI*-digested pCOMPASS31. The resulting plasmid was called Destination vector II. Through this, 45-bp *LHR – PmeI - PciI*- 45-bp *RHR* were introduced into the plasmid to allow integration of Destination vector II, after digestion with either *PmeI* or *PciI*, in *ADE2.a<sup>10</sup>* site of the yeast genome.

### Cloning of parts

*Construction of the ATF/BS library:* We selected three *JUB1*-, two *ANAC102*-, two *ATAF1*-, one *RAV1*-, and one *GRF7*-derived ATFs. Coding sequences of ATFs were obtained by PCR using appropriate expression plasmids<sup>7</sup> as templates and the respective forward (ATF-for) and reverse (ATF-rev) primers. The corresponding binding sites (*JUB1 2X*, *JUB1 4X*, *ANAC102 4X*, *ATAF1 2X*, *RAV1 4X*, and *GRF7 4X*) fused upstream to the yeast minimal *CYC1* promoter were obtained by PCR using appropriate reporter plasmids<sup>7</sup> as templates and the respective forward (BS-for) and reverse (BS-rev) primers. Both fragments were inserted into Entry vector X previously digested with *FseI/AscI*. Constructs containing the ATFs NLS-GAL4AD-RAV1 and

NLS-GAL4AD-GRF7 (both in combination with four copies of their binding sites upstream of the minimal *CYC1* promoter)<sup>7</sup> were cloned into the Entry vector X by standard overlap-based cloning. The remaining ATF/BS combinations were assembled using a combinatorial approach (see **Results**). For JUB1, DNA fragments encoding three JUB1-derived ATFs were mixed in 1:1 molar ratio, and two promoter fragments containing two and four copies of the JUB1 BS, respectively, were mixed in 1:1 molar ratio. Overlap-based cloning results in different combinations between the three ATFs and the two binding sites in the Entry vector, including the three desired combinations. For ATAF1, DNA fragments encoding two ATAF1-derived ATFs were mixed in 1:1 molar ratio and assembled by overlap-based cloning with a promoter fragment containing two copies of the ATAF1 binding site. For ANAC102, DNA fragments encoding two ANAC102-derived ATFs were mixed in 1:1 molar ratio and assembled by overlap-based cloning with a promoter fragment containing four copies of the ANAC102 binding site. In this way, nine different Entry vectors X derivatives harboring the different ATF/BS regulator modules were generated. The desired constructs were identified by colony PCR followed by sequencing (ATF: primers EXSEQ-for and EXSEQ-rev; *BS-Pro<sub>CYC1\_mini</sub>* and *Pro<sub>TDH3</sub>*: PROSEQ-for and PROSEQ-rev). Primer sequences are given in **Supplementary Table 9**.

#### *Assembly of pathway genes*

$\beta$ -Carotene and  $\beta$ -ionone: The promoters of the *E. coli* selection marker genes, yeast terminators, relevant parts of the CDSs of the  $\beta$ -carotene (*BTS1* or *GGPPSbc*, codon optimized *Mcr1* and *Mcr1YB*) and  $\beta$ -ionone (*RiCCD1*) biosynthesis genes were amplified from different sources of genomic DNA or plasmids. Entry vector X was digested with *FseI* and *AscI* and three fragments including the CDS of the gene of interest, the yeast terminator and the promoter of the *E. coli* selection marker were inserted. The primers used for amplification included overhangs with rare restriction enzyme recognition sites compatible to the appropriate Acceptor vector for the next cloning steps. All parts were verified by sequencing.

Naringenin: The promoters of the *E. coli* selection marker genes, relevant parts of the CDSs of the NG biosynthetic pathway (*AtC4H:L5:AtATR2*, *PhCHI*, *HaCHS*, *At4CL-2*, and *AtPAL-2*) fused to the terminators were amplified from different sources of plasmids and mixed in equimolar ratio.<sup>11</sup> The Acceptor vectors were digested with *FseI* and *AscI* (Acceptor vector E, F, G, and H) or *XhoI* and *BamHI* (Destination vector II) and mixed in equimolar ratio. The combined fragments containing the CDSs, promoters of the *E. coli* selection marker genes and the digested plasmids were mixed in a single tube in a ratio recommend by the manufacturer to perform the NEBuilder HiFi reaction. We plated the transformed cells onto four different LB agar media containing either



spectinomycin, ampicillin, chloramphenicol, or triclosan as selection markers for the transformed plasmids. Thereby, *AtC4H::L5::AtATR2*, *PhCHI*, *HaCHS*, *At4CL-2*, and *AtPAL-2* were assembled in Destination vector II, Acceptor vector E, F, G, and H, respectively.

All parts, their functions, and sources are given in **Supplementary Table 3**.

#### *Combinatorial expression of each single gene of pathways*

$\beta$ -Carotene and  $\beta$ -ionone: Nine PCR-amplified ATF/BS fragments (primers X0\_for and Z0\_rev, PCR performed on the Entry vectors-nine ATF/BS) were mixed in equimolar amounts. The *McrtI*, *BTS1*, *McrtYB*, and *RiCCD1* coding sequences and their downstream terminators and the promoters of the *E. coli* selection marker genes were PCR-amplified from Entry vectors using appropriate pairs of primers and mixed in equimolar amounts. Four vectors were digested: Destination vector I (*Sall/EcoRI*) and Acceptor vector A to C (*Fsel/Ascl*), and mixed in equimolar ratio. The combined ATF/BS fragments, the fragments containing the CDSs, and the digested plasmids were mixed in a single tube in a ratio recommend by the manufacturer to perform the NEBuilder HiFi reaction. We plated the transformed cells onto four different LB agar medium containing either spectinomycin, ampicillin, chloramphenicol, or triclosan as selection markers for the transformed plasmids. Cells harboring the successfully assembled constructs are able to grow and form colonies on appropriate selection medium. Hence, *McrtI*, *BTS1*, *McrtYB*, and *RiCCD1* were assembled in Destination vector I and Acceptor vectors A, B, and C, respectively. Moreover, the *GGPPSbc* coding sequence and its downstream terminator and the promoters of the *Amp<sup>R</sup>* selection marker were PCR-amplified from Entry vector X using appropriate pairs of primers and mixed in equimolar amounts. The combined ATF/BS fragments, the fragment containing the *GGPPSbc* CDS, and the *Fsel/Ascl*-digested Acceptor vector A were mixed in a single tube in a ratio recommend by the manufacturer to perform the NEBuilder HiFi reaction. We plated the transformed cells onto LB agar medium containing ampicillin allowing cells harboring the successfully assembled constructs to grow. The constitutive and strong yeast *TDH3* promoter<sup>14</sup> was used as a positive control in all experiments. PCR-amplified *Pro<sub>TDH3</sub>* (*TDH\_for/TDH\_rev*, on on BY4741 genomic DNA) was cloned in *Fsel/Ascl*-digested Entry vector X (**Supplementary Table 4**).

Naringenin: Nine PCR-amplified ATF/BS fragments were mixed in equimolar amounts. The *AtC4H:L5:AtATR2*, *PhCHI*, *HaCHS*, *At4CL-2*, and *AtPAL-2* coding sequences and their downstream terminators and the promoters of the *E. coli* selection marker genes were PCR-amplified from Destination vector II and Acceptor vectors E - H, respectively, and mixed in equimolar amounts. Five digested vectors, Destination vector II (*XhoI/BamHI*) and Acceptor

vectors E to H (*FseI/AscI*), the *ATF/BS* fragments, and the fragments containing the naringenin biosynthesis CDSs were mixed in a single tube in a ratio recommend by the manufacturer to perform the NEBuilder HiFi reaction. We plated the transformed cells onto five different LB agar media containing either spectinomycin, ampicillin, chloramphenicol, triclosan, or gentamicin. Cells with successfully assembled constructs are able to grow on the appropriate selection medium. Thereby, *AtC4H:L5:AtATR2*, *PhCHI*, *HaCHS*, *At4CL-2*, and *AtPAL-2* were assembled in Destination vector II and Acceptor vectors E, F, G and H, respectively (**Supplementary Table 4**).

### **Combinatorial cloning of $\beta$ -carotene and $\beta$ -ionone pathway genes in Destination vector I**

Equal amounts of the nine Destination vectors I-McrtI were mixed and digested with *I-CeuI/SbfI*. Equal amounts of the nine Acceptor vectors A-BTS1 were mixed and used as a template to amplify fragments containing the promoter of the *HIS3* auxotrophic marker, the IPTG-inducible promoter, the *ATF/BS* fragments, the *BTS1* CDS, the *TDH3* terminator, the promoter and CDS of the ampicillin selection marker using respective forward (X0\_for) and reverse (Z2\_rev) primers. Therefore, PCR-amplified fragments contain nine different modules differing in their *ATF/BS* units. Using TAR, the PCR-amplified fragments were inserted into the nine Destination vectors I-McrtI to generate a library of  $9^2 = 81$  Destination vectors I-McrtI-BTS1. Yeast cells with successful constructs grow on SC-Ura/-His medium. The plasmid library was recovered from the yeast cells, transformed into *E. coli*, and grown on LB agar plates containing ampicillin, where only cells harboring correct assemblies will grow. In the next step, the plasmid library was digested with *FseI/I-SceI*. Equal amounts of the nine Acceptor vectors B-McrtYB were mixed and used as a template for a PCR reaction to amplify fragments containing the promoter of the *LEU2* auxotrophic marker, the IPTG-inducible promoter, the *ATF/BS* fragments, the *MctYB* CDS, yeast synthetic terminator 3, the promoter and CDS of the chloramphenicol selection marker using respective forward (X0\_for) and reverse (Z3\_rev) primers. Therefore, PCR-amplified fragments contain nine different modules differing in their *ATF/BS* units. Using TAR, the PCR-amplified McrtYB modules were inserted into the linearized Destination vector I-McrtI-BTS1 library to generate a new library of  $9^3 = 729$  Destination vectors I-McrtI-BTS1-McrtYB. Yeast cells with successful constructs grow on SC-Ura/-His/-Leu medium. The library of Destination vector I containing all three genes leads to the production of  $\beta$ -carotene in inducing media. The plasmid library was recovered from the yeast cells and transformed into *E. coli*, which was selected on LB agar medium containing chloramphenicol. The recovered library was subsequently used to

integrate the pathway genes including the ATF/BS regulatory modules into the yeast genome (see below).

Plasmids producing the highest amounts of  $\beta$ -carotene in each background strain, IMX672.1 and Gen 0.1, were subsequently used for  $\beta$ -ionone production. The plasmids were digested with *Pi*-*PspI*. Equal amounts of the nine Acceptor vectors C-RiCCD1 were mixed and used as a template to amplify fragments containing the promoter of the *TRP1* auxotrophic marker, the IPTG-inducible promoter, the ATF/BS regulatory units, the *RiCCD1* CDS, the yeast *TEF1* terminator, and the promoter and CDS of the triclosan selection marker using respective forward (X0\_for) and reverse (Z4\_rev) primers. Therefore, PCR-amplified fragments contain nine different modules differing in their ATF/BS regulatory modules. Using TAR, the PCR-amplified fragments were inserted into the Destination vector I-McrtI-BTS1-McrtYB. Yeast cells with correctly assembled constructs grow on SC-Ura/-His/-Leu/-Trp medium. The library of Destination vector I containing all four genes leads to the production of  $\beta$ -ionone. The plasmid library was recovered from the yeast cells, transformed into *E. coli*, and grown on LB agar medium containing triclosan. Plasmids pCAROTENE-PTDH3 and pIONONE-PTDH3 (see next chapter) contain the  $\beta$ -carotene and  $\beta$ -ionone CDSs under the control of the *TDH3* promoter assembled in Destination vector I.

### Positive control for $\beta$ -carotene and $\beta$ -ionone production

Destination vector I was digested with *EcoRI/SalI* and used in a two-way assembly reaction using PCR-amplified *Pro*<sub>TDH3</sub> sequence (PROTDHMCRTI\_for/PROTDHMCRTI\_rev, on pCP11<sup>13</sup>) and *McrtI* CDS fused to *Ter*<sub>FBA1</sub> (MCRTIPOS\_for/MCRTIPOS\_rev, on Entry vector X-McrtI) to generate pMCRTI\_PTDH3.

*PciI/ClaI*-digested pCP11<sup>13</sup> was used in an assembly reaction using PCR-amplified *HIS3* encoding sequence (PCPHIS\_for/PCPHIS\_rev, on pGN005B<sup>7</sup>) to generate pCOMPASS32. Next, *NotI*-digested pCOMPASS32 was used in an assembly reaction using *BTS1* CDS (BTSPPOS\_for/BTSPPOS\_rev, on Entry vector X-BTS1) to generate pBTS1\_PTDH3.

*PciI/ClaI*-digested pCP11<sup>13</sup> was used in an assembly reaction using PCR-amplified *LEU2* encoding sequence (PCPLEU\_for/PCPLEU\_rev, on pGAD424, TAKARA Bio, GenBank #U07647) to generate pCOMPASS33. Next, *NotI*-digested pCOMPASS33 was used in an assembly reaction using *McrtYB* CDS (MCRTYBPOS\_for/MCRTYBPOS\_rev, on Entry vector X-McrtYB) to generate pMCRTYB\_PTDH3.

Acceptor vector C was digested with *FseI/AscI* and used in a two-way assembly reaction using PCR-amplified *Pro*<sub>TDH3</sub> sequence (PROTDHRICCD\_for/PROTDHRICCD\_rev, on pCP11<sup>13</sup>) and

*RiCCD1* CDS - *Ter*<sub>TEF2</sub> - *PI-PspI* - *Pro*<sub>TCSR</sub> (RICCDPOS\_for/RICCDPOS\_rev, on Entry vector X-RiCCD1) to generate pRICCD1\_PTDH3.

*I-CeuI*- digested pMCRTI\_PTDH3 was used in an assembly reaction using PCR-amplified *Pro*<sub>TDH3</sub> - *BTS1* - *Ter*<sub>TDH3</sub> (PTDH3\_BTS1\_for/PTDH3\_BTS1\_rev, on pBTS1-PTDH3) and AmpR encoding sequence (AMPR\_for/Y1\_rev, on pGN003B<sup>7</sup>). The resulting plasmid was called pCOMPASS34. Then, *FseI/I-SceI*-digested pCOMPASS33 was used in an assembly reaction using PCR-amplified *Pro*<sub>TDH3</sub> - *McrTYB* (PTDH3\_MCRTYB\_for/PTDH3\_MCRTYB\_rev, on pMCRTYB-PTDH3) and *Ter*<sub>SYN3</sub> - *CmR* - *Ter*<sub>CmR</sub> (CMR\_for/Y2\_rev, on pLD\_3\_4<sup>13</sup>). The resulting plasmid was called pCAROTENE-PTDH3. Then, *PI-PspI*-digested pCOMPASS34 was used in an assembly reaction using PCR-amplified *Pro*<sub>TDH3</sub> - *RiCCD1* - *Ter*<sub>TEF2</sub> - *TCSR* - *Ter*<sub>TCSR</sub> (PTDH3\_RICCD1\_for/Y3\_rev, on pRICCD1\_PTDH3.). The resulting plasmid was called pIONONE-PTDH3.

### Integration of Destination vectors into the yeast genome

Strains IMX672.1 or Gen 0.1 were transformed with the library of *Bam*HI- or *NotI*-linearized Destination vectors I resulting in cassette integration into the yeast *LYS2* or *URA3* loci, respectively, thereby leading to the integration of the  $\beta$ -carotene or  $\beta$ -ionone pathway genes.

### Multi-locus integration of $\beta$ -carotene and $\beta$ -ionone pathway genes

To simultaneously integrate the genes required for  $\beta$ -carotene (*Mcr1I*, *BTS1*, and *McrTYB*) or  $\beta$ -ionone (*Mcr1I*, *BTS1*, *McrTYB* and *RiCCD1*) production, strains Gen 0.1 and IMX672.1 were co-transformed with either 1  $\mu$ g of triple sgRNA plasmid pCOM004<sup>6</sup> (to integrate *Mcr1I*, *BTS1*, and *McrTYB* into the *X-3*, *XI-2*, and *XII-5* locus, respectively) or pCOM006 (to integrate the *RiCCD1* containing module into the *ura3-52* locus), plus 1  $\mu$ g of each donor fragment: *Mcr1I* (primers DES1-X3\_for and DES1-X3\_rev on Destination vector I-Mcr1I), *BTS1* (primers ACCEPTA-XII5\_for and ACCEPTA-XII5\_rev on Acceptor vector A-BTS1), *McrTYB* (primers ACCEPB-XI3\_for and ACCEPB-XI3\_rev on Acceptor vector B-BTS1), and *RiCCD1* (primers ACCEPTC-ura\_for and ACCEPTC-ura\_rev on Acceptor vector C-RiCCD1). To select for  $\beta$ -carotene producing strains, cells were plated on media that selected for the presence of the sgRNA plasmid which harbors the G418 resistance gene. Colonies were washed off the plate(s) and subsequently cells were plated on selective medium (SC-Leu/-Ura/-His) to screen for  $\beta$ -carotene producing colonies with integrated selection markers. Subsequently, when colonies appeared, the transformation plates were replicated on non-selective induction plates (YPDA, 2% (w/v) galactose, 20  $\mu$ M IPTG). To select for  $\beta$ -ionone pathway strains, cells were plated on SC-Lys to

select for the presence of plasmid pCOM006. Colonies were washed off and plated on selective medium to screen for  $\beta$ -ionone producing colonies with integrated selection markers. When colonies appeared, the transformation plates were replicated on non-selective induction plates. Multiplex PCR was performed on single colonies; primers were designed to amplify ATF/BS regulatory units upstream of each CDS.

### **Multi-locus integration of naringenin and $\beta$ -ionone pathway genes in multiple loci of yeast strain MXFde 0.2**

Strain MXFde 0.2 was co-transformed with pCOM006 and 1  $\mu$ g of *RiCCD1* donor (primers ACCEPTC-ura\_for and ACCEPTC-ura\_rev on Acceptor vector C-RiCCD1) to integrate the *RiCCD1* containing module into the *ura3-52* site. To select for successful integration, cells were plated on SC-Lys/-Trp, as the *LYS2* marker is encoded on the plasmid and the *TRP1* selection marker is present on the integrated donor. When colonies appeared, the transformation plates were replicated on non-selective induction plates. Next, to integrate five genes required for naringenin production, the library of  $\beta$ -ionone producing strains was co-transformed with 1  $\mu$ g of sgRNA plasmid pCOM004 (to integrate *PhCHI* and *HaCHS* into the *X-3* and *XI-2* locus, respectively), pCOM007 (to integrate *AtC4H::L5::AtATR2* into *XII-2*), pCOM008 (to integrate *At4CL-2* and *AtPAL-2* into the *X-2* and *XI-5* locus, respectively), plus 1  $\mu$ g of each donor fragment: *PhCHI* (primers ACCPTE-X3\_for and ACCPTE-X3\_rev on Acceptor vector E-CHI), *HaCHS* (primers ACCPTF-XI2\_for and ACCPTF-XI2\_rev on Acceptor vector F-CHS), *AtC4H::AtATR2* (primers DESII-XII2\_for and DESII-XII2\_rev on Destination vector II-C4H::ATR2), *At4CL-2* (primers ACCPTG-X2\_for and ACCEPTG-X2\_rev on Acceptor vector G-4CL-2), and *AtPAL-2* (primers ACCEPTH-XI5\_for and ACCEPTH-XI5\_rev on Acceptor vector H-PAL2). Cells were plated on media that selected for the presence of the sgRNA plasmids which encodes the *LYS2* selection marker. When colonies appeared, the transformation plates were replicated on selective plates (YPDA with five dominant selection markers) to (i) remove the episomal plasmids and (ii) screen for successfully integrated genes into the genome. To simultaneously integrate the three CDSs (in addition to *RiCCD1*) required for  $\beta$ -ionone production in yeast, the library of naringenin strains was co-transformed with 1  $\mu$ g of triple-sgRNA plasmid pCOM005 (to integrate *Mcrtl*, *McrtyB*, and *BTS1* or *GGPPSbc* into the *X-4*, *XI-3*, and *XII-5*, loci, respectively), plus 1  $\mu$ g of the *Mcrtl*, *McrtyB*, and *BTS1* donor fragments: *Mcrtl* (primers DES1-X4\_for/DES1-X4\_rev on Destination vector I-Mcrtl), *McrtyB* (primers ACCEPB-XI3\_for/ACCEPTB-XI3\_rev on Acceptor vector B-McrtyB), and *BTS1* (primers ACCEPTA-XII5\_for/ACCEPTA-XII5\_rev on Acceptor vector A-BTS1). Moreover, we used the *GGPPSbc*

donor (primers ACCEPTA-XII5\_for/ACCEPTA-XII5\_rev on Acceptor vector A-GGPPSbc) in addition to the *BTS1* donor. Cells were plated on media that selected for the presence of the sgRNA plasmids and successfully integrated donors (SC-Ura/-Leu/-His/-Lys). Subsequently, when colonies appeared, the cultures of the transformation plates were replicated on non-selective induction plates. When colonies appeared, they were scraped of the plates and used to establish liquid-media cultures thereby generating a library of strains (called Narions) producing  $\beta$ -ionone and naringenin at different levels.

### **Induction experiments**

Yeast strains harboring the pathway genes were plated on SC (synthetic complete) medium on non-inducing medium (2% (w/v) glucose) with appropriate selection markers. Cells were grown at 30°C for 3 - 4 d. The plates were scraped to collect the cells which were then plated on induction medium plates containing 20 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG), 2% (w/v) galactose, 1% (w/v) raffinose and the appropriate selection markers (inducing and selective media). Cells were grown at 30°C for 3 - 4 d. Episomal plasmids were isolated from the collected colonies, transformed into *E. coli* and the assembled parts were sequenced. In the case in genome-integrated cassettes, colony PCR was performed, followed by sequencing. Three independent colonies from either integrated or episomal constructs introduced into the Gen 0.1 or IMX672 backgrounds were chosen for product ( $\beta$ -carotene or  $\beta$ -ionone) analysis by HPLC. To this end, colonies were selected based on intense orange color, indicating high level of  $\beta$ -carotene level, or light orange color, indicating accumulation of  $\beta$ -ionone. Positive controls expressed the pathway genes under the control of the yeast *TDH3* promoter.

### **Recovery of plasmids from yeast**

Assembled constructs encoding the  $\beta$ -carotene and  $\beta$ -ionone pathways were recovered from yeast as previously described<sup>15</sup>. All colonies were scraped; constructs were recovered of yeast cultured on media with appropriate yeast selection markers. Recovered plasmids were transformed into *E. coli* cells and plated on LB plates with appropriate *E. coli* selection markers.

### **High-performance liquid chromatography (HPLC)**

Single colonies of yeast strains were inoculated into 4 ml non-inducing SC medium with appropriate selection marker (pre-culture), and grown for 18 - 24 h at 30°C in a rotary shaker at 230 rpm. The pre-cultures were then used to inoculate main cultures (50 ml) inducing SC medium (20 mM IPTG, 2% galactose) with the appropriate selection marker. All flask cultures

were inoculated from pre-cultures grown on the same medium, to an initial OD<sub>600</sub> of 0.1. Cells were grown in 500-ml flask at 30°C for 3 d in a rotary shaker at 230 rpm to saturation and then harvested for HPLC analysis of metabolites. Carotenoid extraction was carried out from cellular pellets according to the acetone extraction method<sup>16</sup>, with some modifications as follows. The cell pellet was washed once with deionized water. Glass beads (400 - 600 µm diameter; Sigma Chemical Co.) and 500 µl of acetone were added to the cell pellet in a 2.0-ml microcentrifuge tube (*Eppendorf*), and vortexed to pulverize the cells, followed by incubation for 20 min at 30°C. After breakage, the bead-cell mixture was centrifuged in a table-top micro-centrifuge at 13,000 rpm for 5 min at 4°C, and the acetone supernatant was collected in standard 2.0-ml *microcentrifuge tubes* (*Eppendorf*). This extraction procedure was repeated until the cell pellet was white. For β-carotene samples, the combined acetone extracts were transferred to a glass vial and dried *using a speed-vac*. To prevent degradation of the carotenoids by light, the glass vials were kept inside a black colored 1.5-ml microcentrifuge tubes (*Eppendorf*). For the determination of β-ionone, half of the combined acetone extract was dried to quantify β-carotene as described above. The other half was kept inside a pharma glass inlay and was used for β-ionone measurement. All operations were carried out under green save light to avoid degradation of carotenoids. The samples were stored at -80°C until further use. Extracts were obtained from cells grown in three independent experiments, and HPLC analyses were performed by AppliChrom (Oranienburg, Germany). The dry samples were solved in acetone before measurement. Carotenoids were separated by HPLC using a RP-HPLC phase AppliChrom OTU DiViDo (250 x 4.6 mm) column with porous 5-µm particles (85/10/5, acetone/methanol/isopropanol, v/v/v) as the mobile phase, with a 1.0-ml/min flux. The elution profiles were recorded using Shimadzu 450 nm (D2) and Kontron 300 nm (D2) detectors.

### **Flow cytometry analysis of naringenin producing cells**

To quantify the yEGFP fluorescence output in the absence of plant-derived ATFs, single colonies of FdeR-based reporter strains were inoculated into 500 µl non-inducing YPDA medium in 48-well deep-well plates. Plates were incubated for 24 h at 30°C and 230 rpm in a rotary shaker. The precultures were used to inoculate main cultures in 500 µl inducing YPDA medium to an OD<sub>600</sub> ~0.1. Cells were grown for 16 h in a rotary shaker at 30°C and 230 rpm. Protein production was inhibited by adding 500 µg/ml cycloheximide. Thereafter, fluorescence output of each cell was analyzed by flow cytometry as described previously.<sup>7</sup>

### **References**

1. Zhang, Y., Werling, U. & Edlmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res* **40**, e55 (2012).
2. Messerschmidt, K., Hochrein, L., Dehm, D., Schulz, K. & Mueller-Roeber, B. Characterizing seamless ligation cloning extract for synthetic biological applications. *Anal Biochem* **509**, 24-32 (2016).
3. Lopez, J. *et al.* Production of beta-ionone by combined expression of carotenogenic and plant CCD1 genes in *Saccharomyces cerevisiae*. *Microb Cell Fact* **14**, 84 (2015).
4. Gietz, R.D. & Schiestl, R.H. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**, 1-4 (2007).
5. Noskov, V.N. *et al.* Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. *Nat. Protocols* **6**, 89-96 (2010).
6. Ronda, C. *et al.* CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microb Cell Fact* **14**, 97 (2015).
7. Naseri, G. *et al.* Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth Biol*, 6,1742-1756 (2017).
8. Lindstrom, D.L. & Gottschling, D.E. The mother enrichment program: a genetic system for facile replicative life span analysis in *Saccharomyces cerevisiae*. *Genetics* **183**, 413-422, 411SI-413SI (2009).
9. Scalcinati, G. *et al.* Combined metabolic engineering of precursor and co-factor supply to increase alpha-santalene production by *Saccharomyces cerevisiae*. *Microb Cell Fact* **11**, 117 (2012).
10. Bao, Z. *et al.* Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS Synth Biol* **4**, 585-594 (2015).
11. Skjoedt, M.L. *et al.* Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nat Chem Biol* **12**, 951-958 (2016).
12. Karpinski, J. *et al.* Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nat Biotechnol* **34**, 401-409 (2016).
13. Hochrein, L. *et al.* AssemblX: a user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Res* (2017).
14. Lee, M.E., DeLoache, W.C., Cervantes, B. & Dueber, J.E. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth Biol* **4**, 975-986 (2015).
15. Noskov, V.N. *et al.* Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. *Nat Protoc* **6**, 89-96 (2011).
16. G. An, D.B. Schuman & E.A. Johnson Isolation of *Phaffia rhodozyma* mutants with increased astaxanthin content. *Appl Environ Microbiol* **55**, 116–124 (1989).

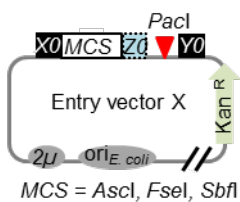


**Supplementary information**

| Color | ATF/BS                             | yEGFP geometric mean |
|-------|------------------------------------|----------------------|
|       | NLS-JUB1-GAL4AD/2X                 | 4043.4 ± 143.3       |
|       | NLS-ATAF1-GAL4AD/2X                | 3612.5 ± 86.3        |
|       | NLS-GAL4AD-ATAF1/2X                | 2568.1 ± 136.4       |
|       | NLS-JUB1-EDLLAD-EDLLAD/4X          | 1890.8 ± 0.2         |
|       | NLS-GAL4AD-ANAC102/4X              | 1616.2 ± 55.9        |
|       | NLS-GAL4AD-GRF7/4X                 | 1138.4 ± 32.4        |
|       | ANAC102-NLS-VP64AD/4X              | 665.7 ± 14.9         |
|       | NLS-DBD <sub>JUB1</sub> -GAL4AD/2X | 479.9 ± 25.4         |
|       | NLS-GAL4AD-RAV1/4X                 | 307.3 ± 18.7         |

**Supplementary Fig. 1. Plant-derived ATF/BSs used for COMPASS.**

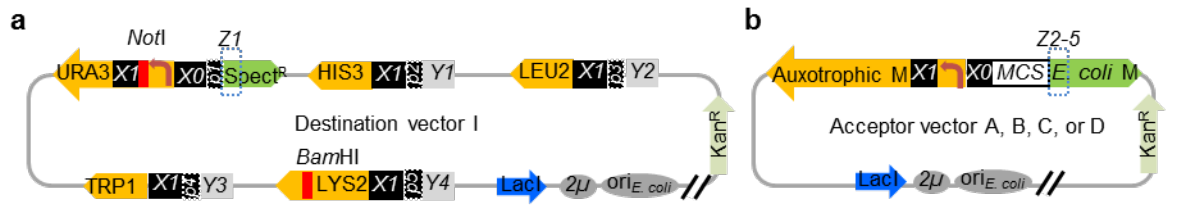
Nine different plant-derived ATFs and promoter pair combinations, providing weak (NLS-GAL4AD-RAV1/4X, NLS-DBD<sub>JUB1</sub>-GAL4AD/2X, and ANAC102-NLS-VP64AD/4X; 300 - 700 AU), medium (NLS-GAL4AD-GRF7/4X, NLS-GAL4AD-ANAC102/4X and NLS-JUB1-EDLLAD-EDLLAD/4X; 1,100 - 1,900 AU) and strong (NLS-GAL4AD-ATAF1/2X, NLS-ATAF1-GAL4AD/2X and NLS-JUB1-GAL4AD/2X; 2,500 - 4,000 AU) transcriptional outputs, were selected from the previously characterized library of plant-derived ATF/BS<sup>7</sup>. AU, arbitrary units, determined using EGFP as reporter<sup>7</sup>. '2X' and '4X' indicate the number of bindings sites implemented for the ATFs within the *CYC1* minimal promoters. The color code is used for presenting the results in **Figs. 5h and 6b**, and **Supplementary Figs. 9d, 10c and 11d**.



**Supplementary Fig. 2. Design of Entry vector X.**

The backbone contains the *E. coli pUC19* replication origin, the kanamycin resistance gene *nptII*, the yeast *2μ* replication origin, an *MCS* flanked by *X0* and *Z0* sequences, and a *Pacl* site flanked by *Z0* and *Y0* sequences. The vector is used to establish ATF/BS units (ATF and BS) within the *MCS*, and to assemble CDS units (CDS, yeast terminator, and the promoter of an *E. coli* selection marker) at the *Pacl* site. *X0*: upstream (left) HR of the vector that provides homology to the forward primer of the first part of the ATF/BS unit (primer ATF\_for). *Z0*: the last 30 bp of the minimal *CYC1* promoter. *Y0*: downstream (right) HR of the vector that provides

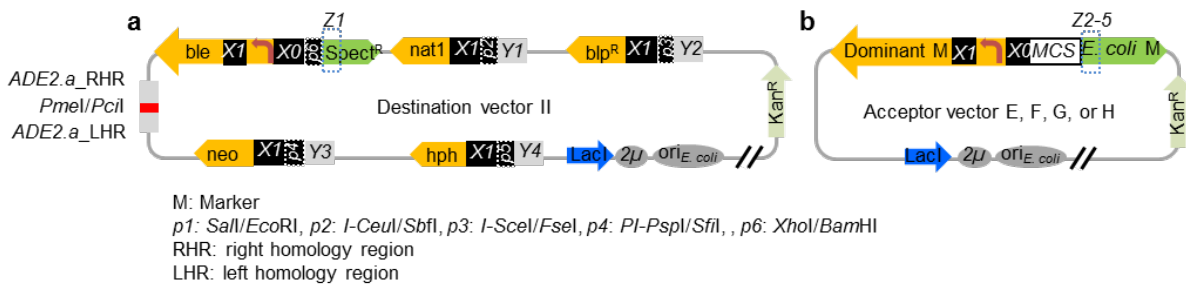
homology to the reverse primers of the last part of the CDS unit (promoter of an *E. coli* selection marker).



M: Marker  
 p1: *SalI/EcoRI*, p2: *I-CeuI/SbfI*, p3: *I-SceI/FseI*, p4: *PI-PspI/SfiI*, p5: *AscI*

**Supplementary Fig. 3. Design of Set 1 vectors.**

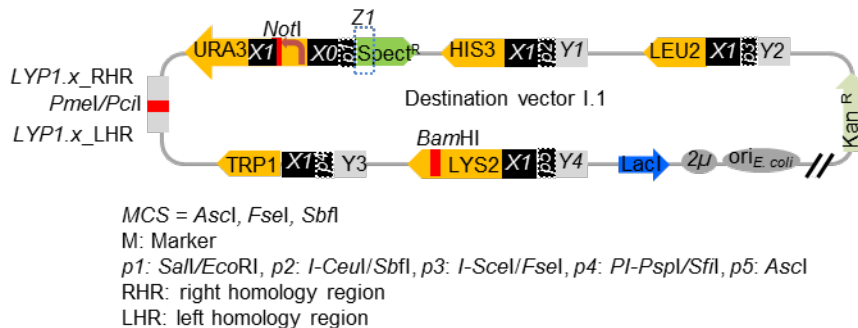
All vectors contain the *E. coli pUC19* replication origin, the kanamycin resistance gene *nptII*, and the yeast  $2\mu$  replication origin. **(a)** Destination vector I has cloning sites p1 (*SalI/EcoRI*), p2 (*I-CeuI/SbfI*), p3 (*I-SceI/FseI*), p4 (*PI-PspI/SfiI*), and p5 (*AscI*). The p1 site is flanked by X0 and Z1, where a *URA3* marker is placed upstream (left) of X0. X0 gives a HR to the Entry vector X, while Z1 gives a HR to the 3' end of *E. coli* markers. The p2, p3, p4, and p5 sites are flanked upstream (left) by X1 and downstream (right) by Y1, Y2, Y3, and Y4, respectively. The CDSs and terminators of *HIS3*, *LEU2*, *TRP1* and *LYS2* are fused upstream (right) to X1 of p2, p3, p4, and p5, respectively. X1 provides a HR with the 5' end of the ATF/BS unit, while Y1, Y2, Y3, and Y4 provide HRs to the last 30 bp of the terminator of *Amp<sup>R</sup>*, *Cm<sup>R</sup>*, *TCS<sup>R</sup>*, and *Gen<sup>R</sup>*, respectively. Destination vector I is equipped with a *NotI* site allowing integration of the plasmid into yeast's *LYS2* locus. **(b)** Acceptor vectors with auxotrophic markers. Acceptor vector A, B, C, or D has an *MCS* flanked by X0 and Z2, Z3, Z4, or Z5, respectively, providing HRs to the appropriate *E. coli* markers. In each Acceptor vector, a yeast auxotrophic marker (A: *HIS3*; B: *LEU2*; C: *TRP1*; and D: *LYS2*) is placed upstream (left) of X0. The HRs X0, X1 and Z1 - Z5 are explained in footnote to **Supplementary Table 2**.



M: Marker  
 p1: *SalI/EcoRI*, p2: *I-CeuI/SbfI*, p3: *I-SceI/FseI*, p4: *PI-PspI/SfiI*, p6: *XhoI/BamHI*  
 RHR: right homology region  
 LHR: left homology region

**Supplementary Fig. 4. Design of Set 2 vectors.**

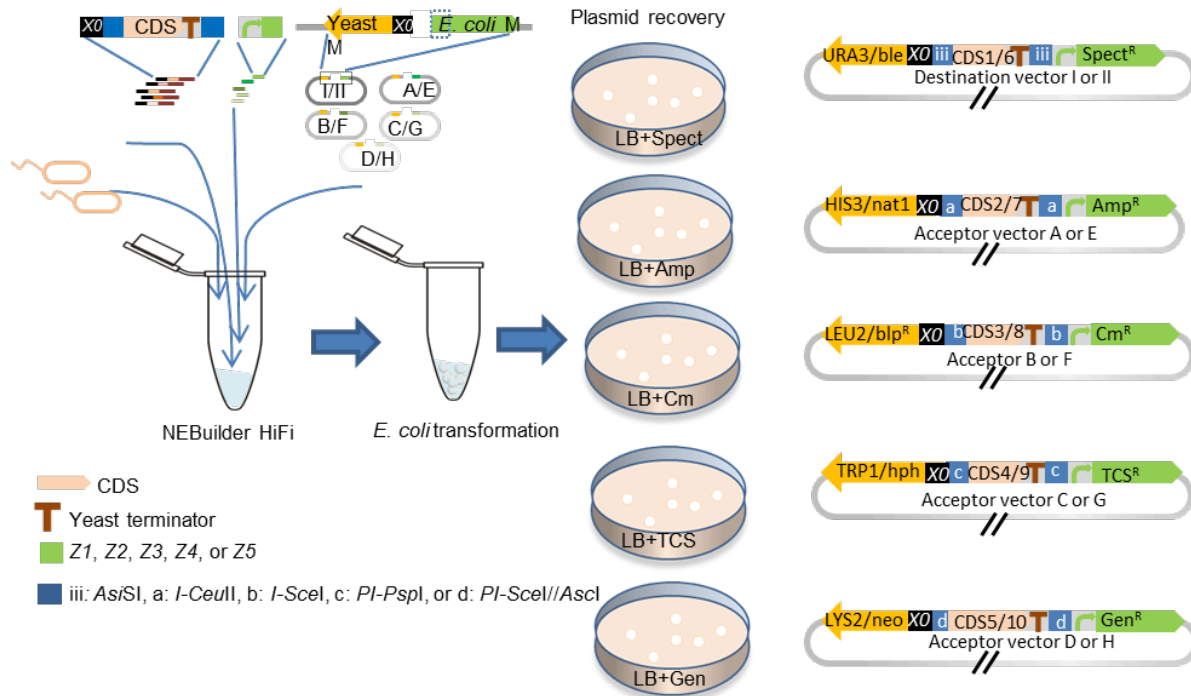
(a) Destination vector II is similar to Destination vector I (see **Supplementary Fig. 3a**), except for the following: (1) site *p5* is replaced by site *p6* (*XhoI/BamHI*), (2) the *URA3* encoding fragment is replaced by a functional *Ble* auxotrophic marker, and (3) the CDSs and terminators of *HIS3*, *LEU2*, *TRP1* and *LYS2* are replaced by the CDSs and terminators of *nat1*, *blp<sup>R</sup>*, *neo*, and *hph*, respectively. Destination vector II is equipped with *PmeI* and *PciI* restriction sites flanked by HRs to allow integration into the *ADE2.a* locus of yeast. (b) Acceptor vectors with dominant markers. Acceptor vectors E, F, G, or H have an *MCS* flanked by an *X0* sequence in the 5' region and *Z2*, *Z3*, *Z4*, or *Z5* sequences in the 3' region. *Z2*, *Z3*, *Z4*, and *Z5* sequences represent the first 30 bp of the *Amp<sup>R</sup>* (E), *Cm<sup>R</sup>* (F), *TCS<sup>R</sup>* (G) or *Gen<sup>R</sup>* (H) CDS, respectively, and provide HRs to the appropriate *E. coli* selection markers present in the CDS units in Level 0. In each Acceptor vector a functional yeast dominant selection marker (E: *nat1*, F: *blp<sup>R</sup>*, G: *neo*, and H: *hph*) is placed upstream (left) of the *X0* sequence. All Destination and Acceptor vectors harbor the *E. coli pUC19* replication origin, the kanamycin resistance gene *nptII*, the yeast *2μ* replication origin, and a yeast selection marker.



**Supplementary Fig. 5. Design of Destination vector I.1.**

Destination vector I.1 contains the *E. coli pUC19* replication origin, the kanamycin resistance gene *nptII*, the yeast *2μ* replication origin, and a yeast selection marker. Moreover, it has cloning sites *p1* (*SalI/EcoRI*), *p2* (*I-CeuI/SbfI*), *p3* (*I-SceI/FseI*), *p4* (*PI-PspI/SfiI*), and *p5* (*AscI*). Site *p1* is flanked by *X0* and *Z1*, and the *URA3* auxotrophic selection marker is located upstream (left) of *X0*. *X0* provides a HR to the Entry vector X, while *Z1* overlaps to the first 30 bp of the *Spect<sup>R</sup>* CDS present in the backbone and provides HRs to the appropriate *E. coli* selection markers present in the CDS units in Level 0. Sites *p2*, *p3*, *p4*, and *p5* are flanked by *X1* in the 5' region, and by *Y1*, *Y2*, *Y3*, and *Y4* in the 3' region. The CDSs and terminators of *HIS3*, *LEU2*, *TRP1* and *LYS2* fused to *X1* are located upstream (right) of sites *p2*, *p3*, *p4*, and *p5*, respectively. *X1* provides an HR to the forward primer amplifying the ATF/BS-CDS module (primer *X1\_for*) in the Level 1 assembly procedure, while the Level 2 *Y1*, *Y2*, *Y3*, and *Y4* provide HRs to the last 30 bp

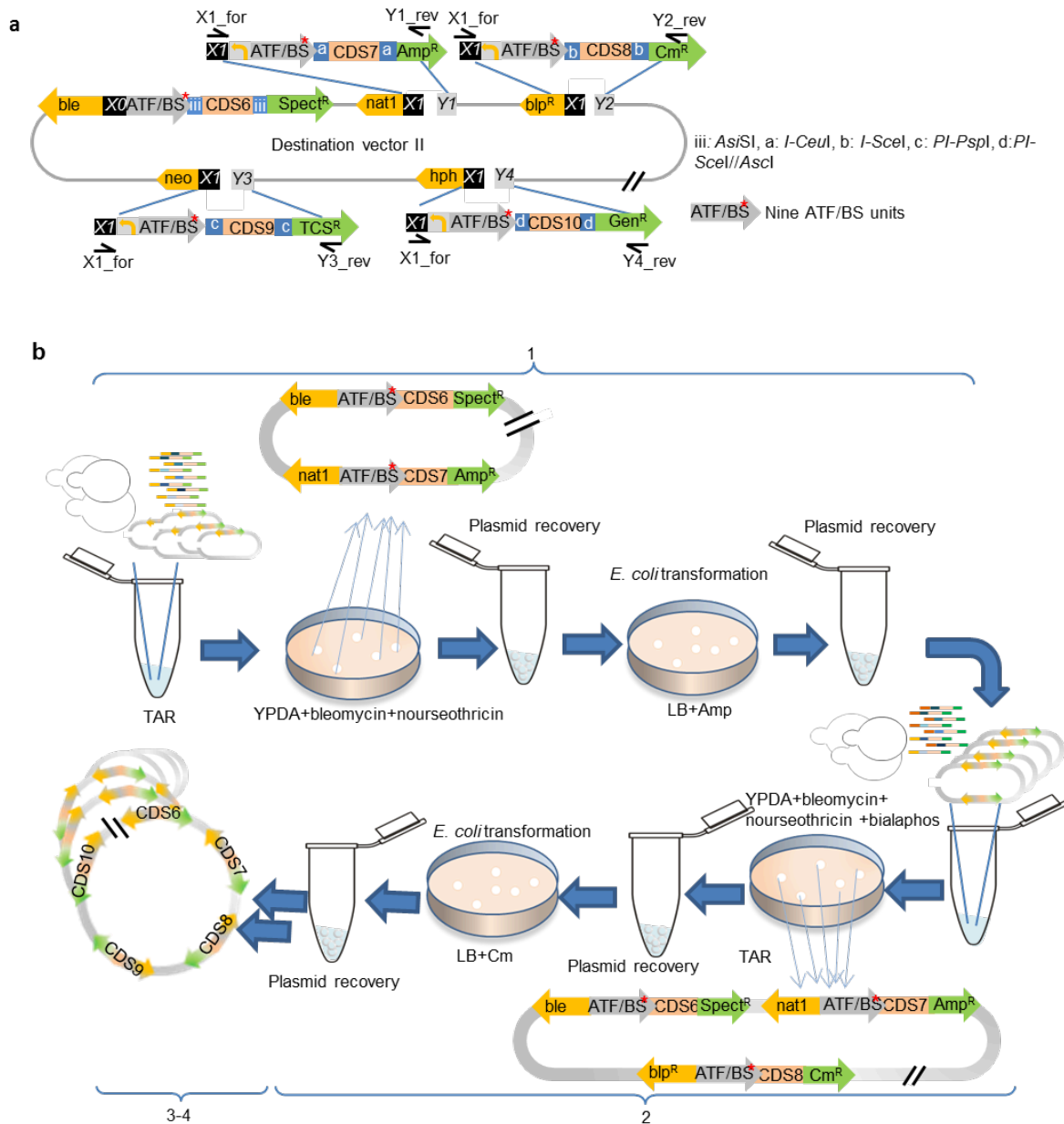
of the terminator sequences of *Amp<sup>R</sup>* (primer Y2\_rev), *Cm<sup>R</sup>* (primer Y3\_rev), *TCS<sup>R</sup>* (primer Y3\_rev), and *Gen<sup>R</sup>* (primer Y4\_rev), respectively. To integrate the pathways assembled in Destination vector I.1 into the genome, it is equipped with *PmeI* (and *PciI*) recognition sequence flanked by HRs for integration into the *LYP1.x* locus<sup>25</sup>.



### Supplementary Fig. 6. Combinatorial assembly of pathway gene units.

Equal amounts of five freely selected CDSs, equal amounts of five yeast terminators, equal amounts of five promoters of *E. coli* selection markers (*Spect<sup>R</sup>*, *Amp<sup>R</sup>*, *Cm<sup>R</sup>*, *TCS<sup>R</sup>*, and *Gen<sup>R</sup>*), and equal amounts of digested vectors from Level 1 (*Sall/EcoRI*-digested Destination vector I or *XhoI/BamHI*-digested Destination vector II, *FseI/Ascl*-digested Acceptor vectors A or E, B or F, C or G, and D or H) are mixed in a single tube for overlap-based cloning. The upstream (left) HR of the CDS-terminator parts overlap with the X0 sequence of the plasmid backbone and rare RE cleavage sites (Destination vectors I or II: iii, Acceptor vectors A or E: a, B or F: b, C or G: c, and E or F: d), compatible to the next level vectors, are introduced before the gene's translation start codon through primer sequences. Moreover, the upstream HR of the promoter of the antibiotic resistance gene is defined by the yeast terminator; the same rare RE recognition site downstream of the terminator is introduced, while the downstream (right) HR of the *E. coli* promoter is defined based on Z sequences (Z1: *Spect<sup>R</sup>*, Z2: *Amp<sup>R</sup>*, Z3: *Cm<sup>R</sup>*, Z4: *TCS<sup>R</sup>*, or Z5: *Gen<sup>R</sup>*). The positive selection of truncated plasmid markers is rendered active by providing promoter sequences during the assembly process. Therefore, Destination vector I-CDS1 or II-

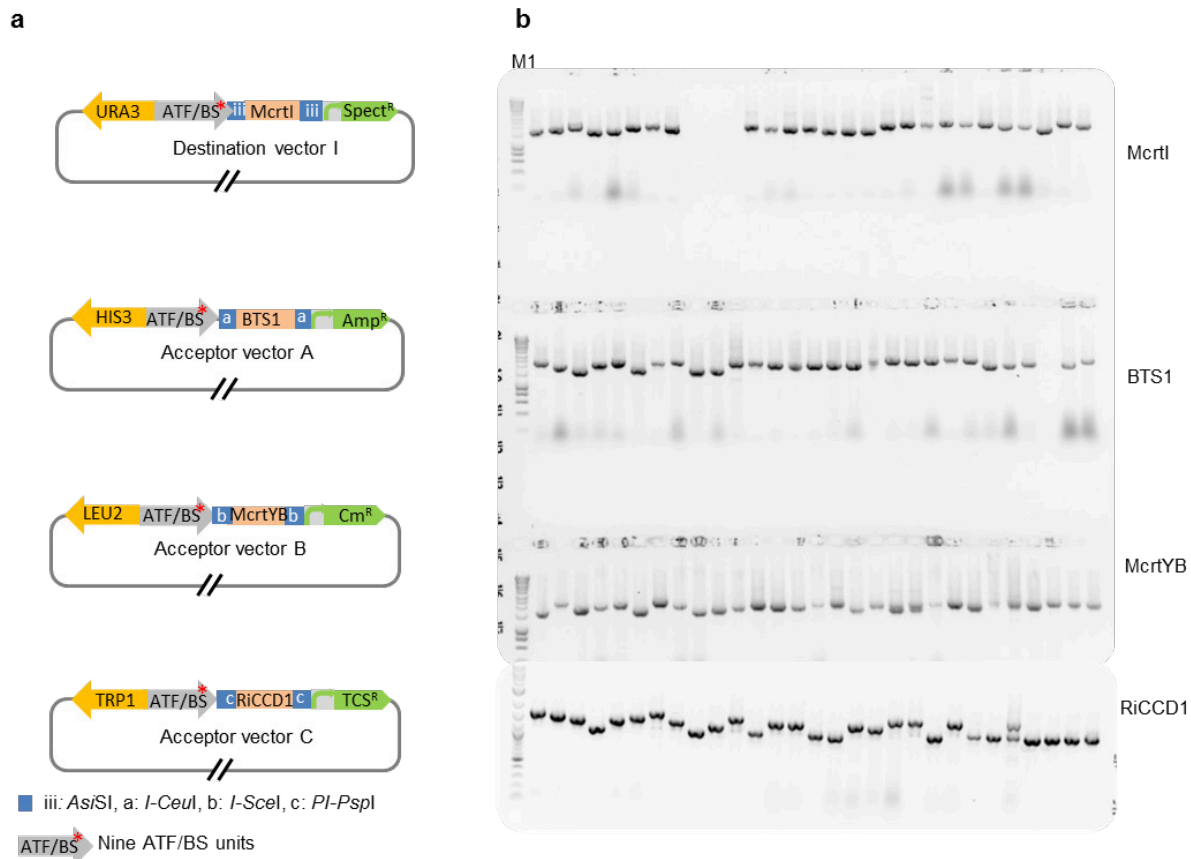
CDS6, Acceptor vectors A-CDS2 or E-CDS7, B-CDS3 or F-CDS8, C-CDS4 or G-CDS9, and D-CDS5 or H-CDS10 are generated. The HRs X0 and Z1 to Z5 are explained in footnote to **Supplementary Table 2**.



**Supplementary Fig. 7. Combinatorial cloning of pathway modules into Destination vector II.**

Libraries of PCR-amplified *Pro<sub>nat1</sub>-ATF/BS-CDS7-Pro<sub>AmpR</sub>-AmpR-Ter<sub>AmpR</sub>* (primers X1\_for/Z2\_rev, on Acceptor vector E), *Pro<sub>blpR</sub>-ATF/BS-CDS8-Pro<sub>CmR</sub>-CmR-Ter<sub>CmR</sub>* (primers X1\_for/Z3\_rev, on Acceptor vector F), *Pro<sub>neo</sub>-ATF/BS-CDS9-Pro<sub>TCSR</sub>-TCSR-Ter<sub>TCSR</sub>* (primers X1\_for/Z4\_rev, on Acceptor vector G), and *Pro<sub>hph</sub>-ATF/BS-CDS10-Pro<sub>GenR</sub>-GenR-Ter<sub>GenR</sub>* (primers X1\_for/Z5\_rev,

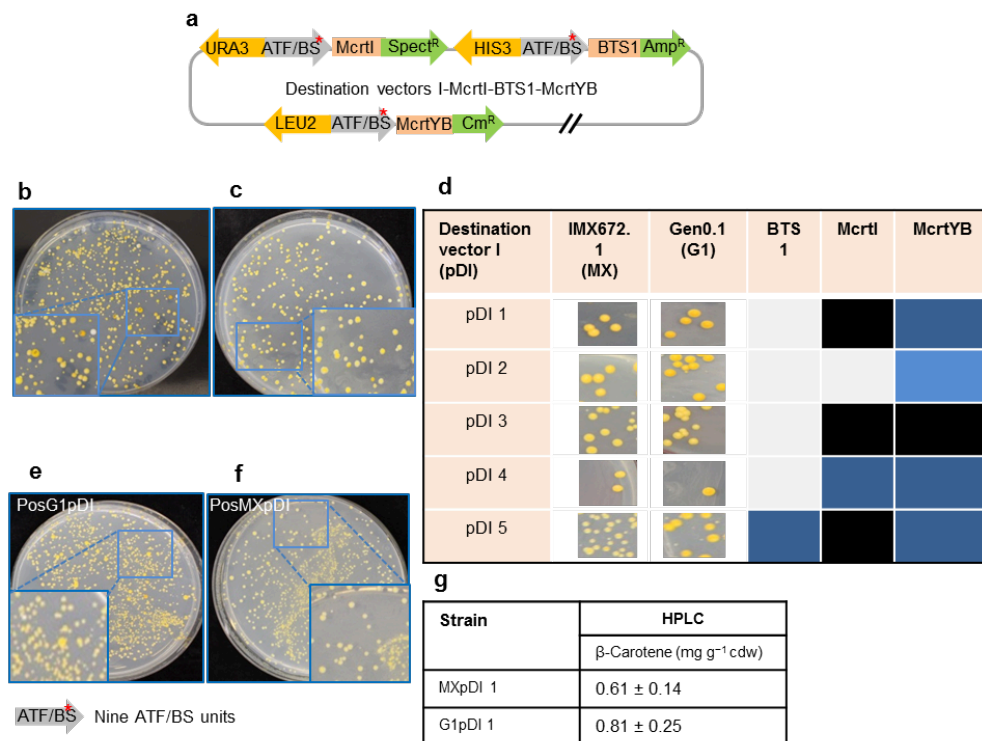
on Acceptor vector H) modules are successively assembled in sites  $p2$ ,  $p3$ ,  $p4$ , and  $p6$ , respectively, starting from the Destination vectors I-CDS6 library (see **Supplementary Fig. 3**) in four rounds of combinatorial cloning. **(b)** The COMPASS workflow for combinatorial assembly of pathway genes in Destination vector II. The mixed ATF/BS-CDS7 modules are assembled in site  $p2$  of Destination vectors II-CDS6 using TAR. Yeast cells with successful assemblies grow on YPDA medium containing bleomycin and nourseothricin. Cells are scraped from the plates and the plasmid library is extracted to obtain a pool of all randomized members. The plasmid library is transformed into *E. coli*, and cells are grown on LB plates containing ampicillin. Next, cells are scraped from the plates and the plasmid library is extracted (1). The ATF/BS-CDS8 modules are assembled in site  $p3$  of Destination vectors I-CDS6-CDS7 using TAR. Yeast cells with successful assemblies grow on YPDA medium containing bleomycin, nourseothricin and bialaphos (2). Using TAR, the libraries of ATF/BS-CDS9 and ATF/BS-CDS10 modules are cloned in sites  $p4$  and  $p6$ , respectively, in two further rounds of combinatorial clonings (3 - 4). For simplicity, IPTG-inducible promoters and terminators are not included in the figure. The HRs  $X0$ ,  $X1$ , and  $Y1$ - $Y4$  are explained in footnote to **Supplementary Table 2**.





### Supplementary Fig. 8. Construction of library of modules for ATF/BS and $\beta$ -ionone pathway genes.

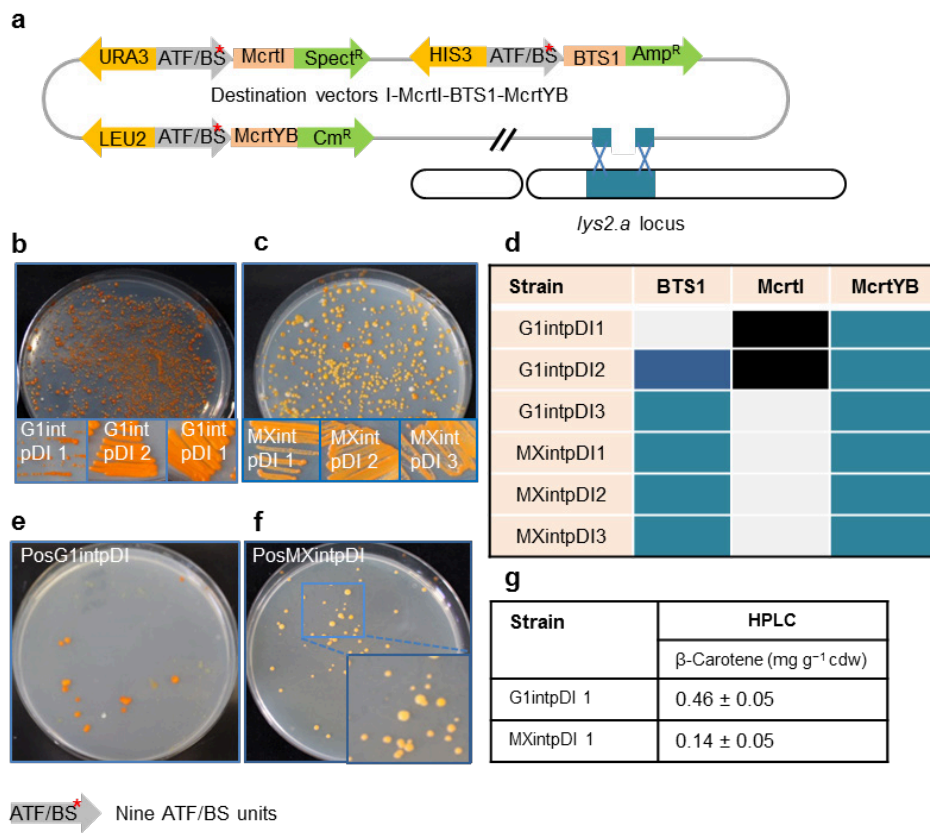
(a) Schematic overview of modules of ATF/BS library and  $\beta$ -ionone pathway genes. Nine ATF/BS control units and the *McrtI*, *BTS1*, *McrtYB*, and *RiCCD1* CDS units of the  $\beta$ -ionone biosynthesis pathway were assembled in Destination vector I, Acceptor vectors A, B, and C, respectively. (b) Agarose gel electrophoresis for the identification of constructed ATF/BS and  $\beta$ -ionone CDS modules. The nine ATF/BS units were assembled upstream of  $\beta$ -ionone CDSs in a single cloning tube in a combinatorial manner. By plating the transformed cells onto LB agar medium containing either spectinomycin, ampicillin, chloramphenicol, or triclosan, libraries of Destination vector I, Acceptor vector A, B, and C, differing in ATF/BS units upstream of *McrtI*, *BTS1*, *McrtYB*, and *RiCCD1* modules, were generated. The verification of transformation was done using colony PCR followed by sequencing (see **Online Methods**). For simplicity, the IPTG-inducible promoters upstream of the ATF/BS, terminator fragments, and HRs needed for cloning are not included in the figure. M1: HyperLadder 1 kb (Bioline).



### Supplementary Fig. 9. $\beta$ -Carotene pathway library in Destination vector I.

(a) Schematic overview of assembly of the  $\beta$ -carotene pathway library in Destination vector I. The libraries of the *McrtI*-, *McrtYB*-, and *BTS1* modules, each under the control of nine ATF/BS

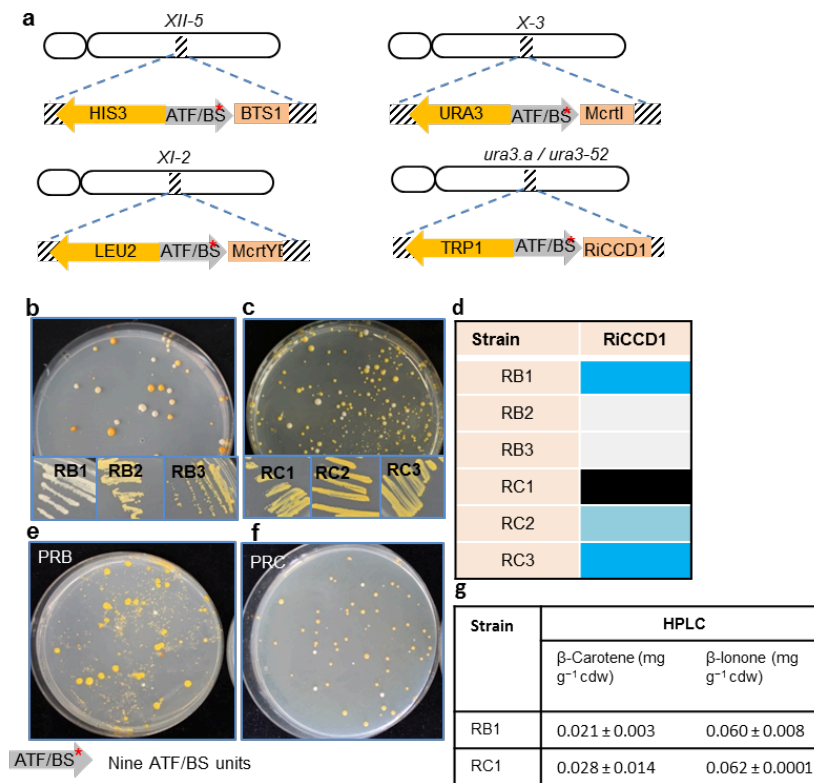
regulators, were assembled into Destination vector I. The ATF/BS-McrtI, -McrtYB-, and -BTS1 modules are flanked by yeast auxotrophic markers *URA3*, *HIS3*, and *LEU2* at the 5' end, while *E. coli* selection markers *Spect<sup>R</sup>*, *Amp<sup>R</sup>*, and *Cm<sup>R</sup>* are placed at the 3' end. Selection on SC-Ura/Leu/His media or LB media containing ampicillin, chloramphenicol, and spectinomycin allows screening for successfully assembled constructs. For simplicity, the IPTG-inducible promoters upstream of the ATF/BS units, terminator fragments, the HRs needed for cloning, the cleavage sites flanking pathway genes and terminators are not included in the figure. **(b)** Transformation of the library of modules containing the  $\beta$ -carotene CDSs under the control of the nine ATF/BS control units into strain Gen 0.1 or **(c)** strain IMX672.1. **(d)** The library of plasmids extracted from yeast was transformed into *E. coli* followed by sequencing. Five sequenced plasmids (pDI 1, pDI 2, pDI 3, pDI 4, and pDI 5) were retransformed into strains Gen 0.1 or IMX672.1. The color code indicating the ATF/BS units assembled upstream of the CDSs is given in **Supplementary Fig. 1**. **(e)** The pCAROTENE-PTDH3 plasmid containing the  $\beta$ -carotene CDSs under the control of the *TDH3* promoter was transformed into strain Gen 0.1 to generate PosG1pDI or **(f)** strain MX672.1 to generate PosMXpDI. **(g)** HPLC analysis of strains MXpDI 1 and G1pDI 1 producing the most intense orange color.





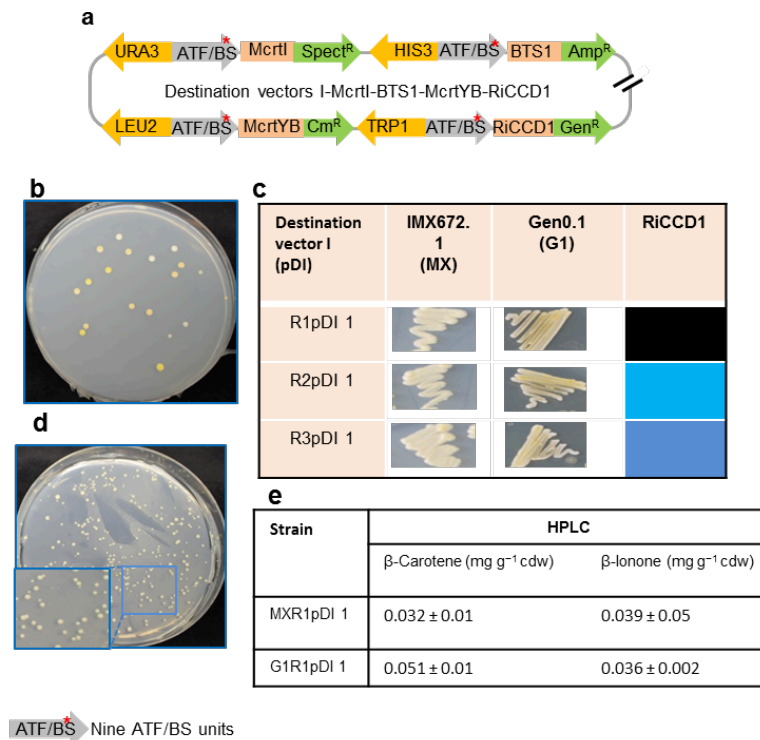
**Supplementary Fig. 10. Single-locus integration of the  $\beta$ -carotene pathway library assembled in Destination vector I.**

(a) Schematic overview of homologous recombination-mediated integration of the Destination vector I encoding the  $\beta$ -carotene pathway into the *lys2.a* locus of yeast, thereby leading to the integration of all pathway genes into a single locus. Selection on SC-Ura/-Leu/-His media allows screening for successfully integrated cassettes. For simplicity, the IPTG-inducible promoters upstream of the ATF/BS, terminator fragments, cleavage sites flanking pathway genes and terminators are not included in the figure. (b) Integration of the library of modules containing the  $\beta$ -carotene CDSs into strain Gen 0.1 or (c) strain IMX672.1. (d) Sequencing results for ATF/BS units present upstream of each CDS from three different colonies producing deep-orange color in Gen 0.1 (G1intpDI 1, G1intpDI 2, and G1intpDI 3) or IMX672.1 (MXintpDI 1, MXintpDI 2, and MXintpDI 3) backgrounds. The color code is given in **Supplementary Fig. 3**. (e) Plasmid pCAROTENE-PTDH3 expressing the  $\beta$ -carotene CDSs under the control of the *TDH3* promoter was integrated into the *lys2.a* locus of strain Gen 0.1 to generate PosG1intpDI or (f) strain IMX672.1 to generate PosMXintpDI. (g) HPLC analysis of strains G1intpDI 1 and MXintpDI 1.



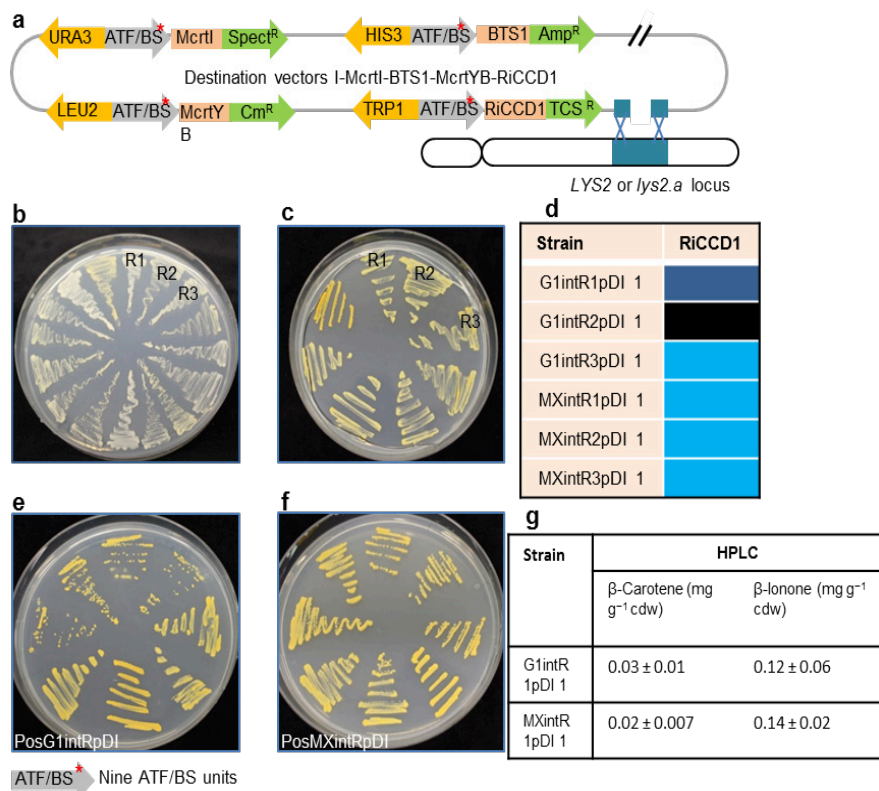
**Supplementary Fig. 11. Multi-locus integration of  $\beta$ -ionone pathway genes.**

**(a)** CRISPR/Cas9-mediated integration of the four  $\beta$ -ionone pathway gene donors into four different loci of the yeast genome. The *McrtI*-, *McrtYB*-, *BTS*- and *RiCCD1*-donors driving  $\beta$ -ionone expression were integrated into the *X-3*, *XI-3*, *XII-5* and *ura3-52* loci. Donors contain the ATF/BS-CDS modules, the yeast dominant markers that are flanked by HRs for integration into the desired loci. Selection on SC-Ura/Leu/His/Trp media allows screening for successfully integrated cassettes. For simplicity, the IPTG-inducible promoters upstream of the ATF/BS, terminator fragments, the HRs needed for cloning, and cleavage sites flanking pathway genes are not included in the figure. **(b)** The library of donors containing the ATF/BS-*RiCCD1* modules was integrated into the *ura3-52* locus of two strains producing high levels of  $\beta$ -carotene, namely strain CB1 with Gen 0.1 background to generate RB variants and **(c)** strain CC1 with IMX672.1 background to generate RC variants. **(d)** Sequencing results for ATF/BS units present upstream of the *RiCCD1* CDS from three different colonies producing light color in either CB1 (RB1, RB2, and RB3) or CC1 (RC1, RC2, and RC3) backgrounds. The color code is given in **Supplementary Fig. 1**. **(e)** Modules containing the  $\beta$ -ionone CDSs under the control of the *TDH3* promoters in strain Gen 0.1 to generate PRB or **(f)** strain IMX672.1 to generate strain PRC. **(g)** HPLC analysis of strains RB1 and RC1.



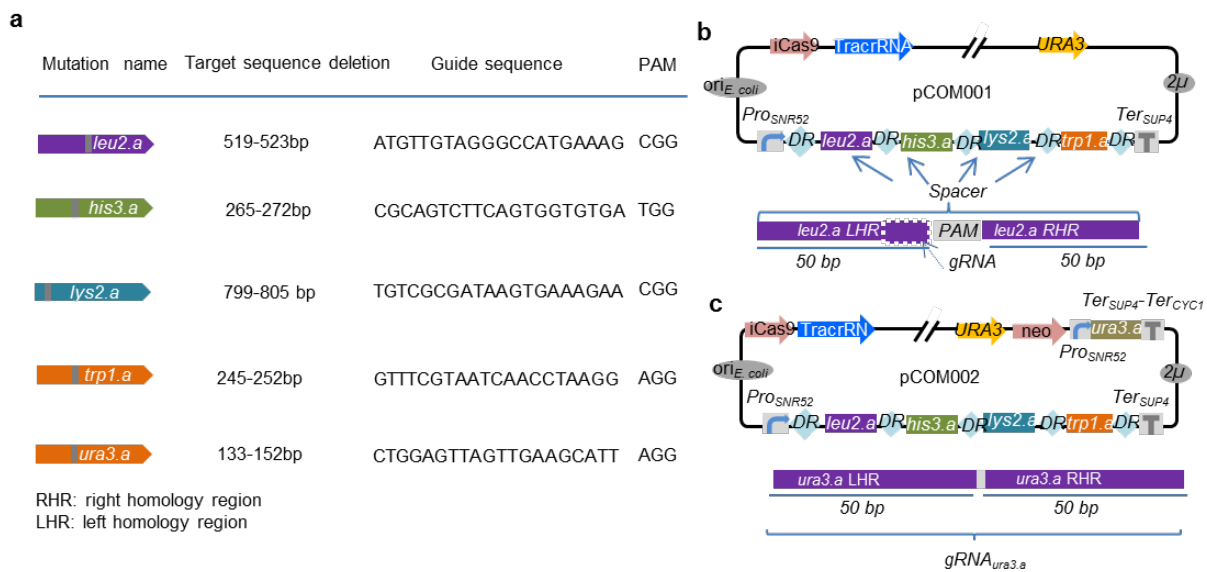
**Supplementary Fig. 12.  $\beta$ -Ionone pathway library in Destination vector I.**

**(a)** Schematic overview of assembly of the ATF/BS and  $\beta$ -ionone pathway library in Destination vector I. The libraries of the *Mcrtl*-, *McrtyB*-, *BTS1*, and *RiCCD1* modules, under the control of ATF/BS regulator units, were assembled into Destination vector I. The ATF/BS-*Mcrtl*-, -*McrtyB*-, -*BTS1*-, -*RiCCD1* modules are flanked by yeast auxotrophic markers *URA3*, *HIS3*, *LEU2*, and *TRP1* at the 5' end, while *E. coli* selection markers *Spect<sup>R</sup>*, *Amp<sup>R</sup>*, *Cm<sup>R</sup>*, and *TCS<sup>R</sup>* are placed at the 3' end. Selection on SC-Ura/-Leu/-His/-Trp media or LB media containing spectinomycin, ampicillin, chloramphenicol, and triclosan allows screening for successfully assembled constructs. For simplicity, the IPTG-inducible promoters upstream of the ATF/BS, terminator fragments, the HRs needed for cloning, the cleavage sites flanking the pathway genes and terminators are not included in the figure. **(b)** Transformation of the library of modules containing the *RiCCD1* CDSs under the control of the nine ATF/BS regulators into strain Gen 0.1. **(c)** Library of plasmids extracted from yeast and transformed into *E. coli* followed by sequencing. Three sequenced plasmids (R1pDI 1, R2pDI 1, and R3pDI 1) were retransformed into Gen 0.1 and IMX672.1 backgrounds. The color code indicating the ATF/BS units assembled upstream of the CDS coding sequences is given in **Supplementary Fig. 1** **(d)** Plasmid pIONONE-PTDH3 containing the  $\beta$ -ionone CDSs under the control of the *TDH3* promoter was transformed into strain Gen 0.1. **(e)** HPLC analysis of MXR1pDI 1 and G1R1pDI 1 producing the shallow yellow color.



**Supplementary Fig. 13. Single-locus integration of the  $\beta$ -ionone pathway library assembled in Destination vector I.**

(a) Schematic overview of double strand break-mediated integration of Destination vector I encoding the  $\beta$ -ionone pathway into the yeast genome. The library of the linearized Destination vector I harboring  $\beta$ -ionone modules ( $\beta$ -ionone pathway genes under the control of the nine ATF/BS units) was integrated into the *lys2.a* locus. Selection on SC-Ura/-Leu/-His/-Trp media allows screening for successfully integrated cassettes. For simplicity, the IPTG-inducible promoters upstream of the ATF/BS, terminator fragments, cleavage sites flanking pathway genes and terminators are not included in the figure. (b) Integration of library of modules containing the  $\beta$ -ionone CDSs into strain Gen 0.1 or (c) strain IMX672.1. (d) Sequencing results for ATF/BS modules present upstream of each CDS from three different colonies producing light color in either Gen 0.1 (G1intR1pDI 1, G1intR2pDI 1, and G1intR3pDI 1) or IMX672.1 (MXintR1pDI 1, MXintR2pDI 1, and MXintR3pDI 1) backgrounds. The color code indicating the ATF/BS units assembled upstream of the CDSs is given in **Supplementary Fig. 1**. (e) pIONONE-PTDH3, a Destination vector I containing the  $\beta$ -carotene CDSs under the control of the *TDH3* promoter, was integrated into strain Gen 0.1 (PosG1intRpDI) or (f) strain IMX672.1 (PosMXintRpDI). (h) HPLC analysis of strains G1intR1pDI 1 and G1intR2pDI 1.



**Supplementary Fig. 14. CRISPR-Cas9-mediated genome modification.**

**(a)** CRISPR-Cas9 target site selection for frame-shift mutations of *LEU2*, *HIS3*, *LYS2*, *TRP1*, and *URA3* genes according to Bao *et al.* (2015)<sup>25</sup>. The relative positions of target sites within each gene and the guide RNA and PAM sequences of each site are shown. **(b)** Schematic presentation of quadruple auxotrophic marker gene disruption using the one-plasmid HI-CRISPR system<sup>25</sup>. Plasmid pCOM001 contains donor sequences for *LEU2*, *HIS3*, *LYS2*, and *TRP1* gene disruption. For each gene disruption, a 100-bp donor sequence containing a gRNA targeting a 20-bp target and the PAM sequence was designed to harbor two 50-bp HRs flanking the Cas9 cleavage site. **(c)** Schematic presentation of quintuple auxotrophic marker gene disruption. Plasmid pCOM002 is derived from pCOM001. It contains a functional yeast *neo* (G418) resistance gene, *Pro*<sub>SNR52</sub> and *Ter*<sub>SUP4</sub> (fused to yeast *Ter*<sub>CYC1</sub>) to control the transcription of the 20-bp gRNA to disrupt *URA3* gene expression, in addition to *LEU2*, *HIS3*, *LYS2*, and *TRP1*. Thereby, transformation of pCOM002 together with the 100-bp *ura3.a* donor into the yeast host cell results in quintuple auxotrophic marker gene disruption.

## Supplementary Notes

### Supplementary Note 1

#### Design of vectors

The Level 0 Entry vector X contains the *E. coli* *pUC19* replication origin, the kanamycin resistance gene *nptII*, the yeast  $2\mu$  replication origin, and a multiple cloning site (MCS) flanked by *X0* and *Z0* sequences (**Supplementary Fig. 2**). A *PacI* site was placed downstream (right) of *Z0*. Entry vector X is used for two types of cloning reactions: to assemble (i) ATF/BS units with diverse (weak, medium and strong) transcriptional outputs, and (ii) CDS units (a yeast terminator, CDS, and the promoter of an *E. coli* selection marker) in a single unit (which defines the next level vector).

Acceptor and Destination vectors harbor the *E. coli* *pUC19* replication origin, the *nptII* kanamycin resistance gene, the yeast  $2\mu$  replication origin, and an MCS flanked with *X0* at the upstream (left) side and different *Z* regions (*Z1* to *Z5*) at the downstream (right) sides. The vectors of Set 1 are equipped with functional auxotrophic selection markers placed upstream (left) of the *X0* region and include Destination vector I (*URA3*) (**Supplementary Fig. 3a**) and Acceptor vectors A (*HIS3*), B (*LEU2*), C (*TRP1*), and D (*LYS2*) (**Supplementary Fig. 3b**). All vectors of Set 2 harbor dominant selection markers upstream (left) of the *X0* region and include Destination vector II (*blp<sup>R</sup>*, bialaphos resistance gene) (**Supplementary Fig. 4a**) and Acceptor vector E (*nat1*, nourseothricin resistance gene), F (*ble*, bleomycin resistance gene), G (*neo*, G418 resistance gene), and H (*hph*, hygromycin B resistance gene) (**Supplementary Fig. 4b**). Moreover, Destination vectors I/II, and Acceptor vectors A/E, B/F, C/G, and D/H harbor the CDSs and terminators of the *Spect<sup>R</sup>*, *Amp<sup>R</sup>*, *Cmr<sup>R</sup>*, *TCS<sup>R</sup>*, and *Gen<sup>R</sup>* genes, respectively, downstream (right) of the cloning sites (MCS in Acceptor vectors; *p1* in Destination vector I, and *p6* in Destination vector II). In addition, Destination vectors harbor *Y1*, *Y2*, *Y3*, and *Y4* sequences representing homology regions (HRs) to the last 30 bp of the terminators of the *Amp<sup>R</sup>*, *Cmr<sup>R</sup>*, *TCS<sup>R</sup>*, and *Gen<sup>R</sup>* genes, downstream (right) of sites *p2*, *p3*, *p4* and *p5*, respectively (**Fig. 2c,e**). Destination vector I and II are equipped with the coding DNA sequence (CDS) and terminators of yeast auxotrophic and dominant selection markers, respectively, upstream (left) of the *X1* region (**Supplementary Fig. 3b,4b**). To integrate the pathways assembled in Destination vectors into the yeast genome, Destination vector I is equipped with *NotI* and *BamHI* restriction sites allowing plasmid integration into the *URA3* or *LYS2* locus, respectively, while *PmeI* and *PciI* recognition sequences flanked by homology arms to integrate into *ADE2.a* locus<sup>25</sup> were added to Destination vector II. Additionally, Destination vector I.1 (**Supplementary Fig. 5**)

contains *PmeI* and *PciI* sequences that are flanked by homology arms for integration into the *LYP1.x* locus<sup>25</sup>.

## Supplementary Note 2

### $\beta$ -Carotene and $\beta$ -ionone production in control strains

As control for  $\beta$ -carotene and  $\beta$ -ionone quantification in COMPASS-derived strains, we measured their production using HPLC for the background strains, namely IMX672.1 and Gen 0.1. For IMX672.1, we did not detect a considerable amount of  $\beta$ -carotene ( $0.007 \pm 0.001$  mg g<sup>-1</sup> cell dry weight (cdw)) or  $\beta$ -ionone ( $0.0008 \pm 0.0005$  mg g<sup>-1</sup> cdw). For Gen 0.1, we did not detect a considerable amount of  $\beta$ -carotene ( $0.015 \pm 0.003$  mg g<sup>-1</sup> cdw) or  $\beta$ -ionone ( $0.0004 \pm 0.0002$  mg g<sup>-1</sup> cdw). Beekwilder *et al.* previously reported  $0.55 \pm 0.005$  mg g<sup>-1</sup> cdw of  $\beta$ -carotene for strain IME167<sup>1</sup>, while it produced  $0.38 \pm 0.08$  mg g<sup>-1</sup> cdw of  $\beta$ -carotene in our experiments (**Supplementary Table 6**) which is 1.3-fold lower than reported. Moreover, they found that strain RiCCD1<sup>1</sup> produced  $0.22 \pm 0.06$  mg g<sup>-1</sup> cdw of  $\beta$ -ionone, while it produced  $0.042 \pm 0.018$  mg g<sup>-1</sup> cdw of  $\beta$ -ionone in our experiments, 5-fold lower than reported.

## Supplementary Note 3

### Construction of a Destination vector library for controllable $\beta$ -carotene production

At Level 2, the library of the ATF/BSs-BTS1 modules was integrated into the library of Destination vector I (pDI)-McrI. Subsequently, the library of ATF/BSs-McrTYB modules was integrated into the library of pDI-McrI-BTS1 (**Supplementary Fig. 9a**), and the combinatorial pDI-McrI-BTS1-McrTYB library was transformed into either strain Gen 0.1 (**Supplementary Fig. 9b**) or IMX672.1 (**Supplementary Fig. 9c**). The selection for successful assemblies was carried out by plating the yeast cells on medium lacking the auxotrophic markers (SC-Ura/Leu/His), as the *URA3*, *LEU2*, and *HIS3* genes are encoded on the assembly modules. A variety of colors ranging from light yellow to deep orange were observed in the different colonies obtained from the Gen 0.1 and IMX672.1 (**Supplementary Fig. 9d**) strains. The colony color may result from the presence of several plasmids within a given yeast cell. The plasmids assembled via COMPASS were recovered from the yeast cells and transformed into *E. coli*. The structure of five plasmids was confirmed by sequencing; the sequenced plasmids were then retransformed into Gen 0.1 and IMX672.1 to confirm functional expression of the pathway. Our results showed that combining a weak ATF/BS (NLS-DBD<sub>JUB1</sub>-GAL4AD/2X) with two strong ATF/BSs (NLS-JUB1-GAL4AD/2X and NLS-ATAF1-GAL4AD/2X) results in superior  $\beta$ -carotene accumulation in both strains (MXpDI 1 and G1pDI 1; **Supplementary Fig. 9d**). Moreover, we observed that

expressing all three pathway genes from strong ATF/BSs (NLS-ATAF1-GAL4AD/2X upstream of *BTS1* and *McrTYB*; and NLS-JUB1-GAL4AD/2X upstream of *McrTI*) in a Destination vector (pDI 5; **Supplementary Fig. 9d**) did not result in high  $\beta$ -carotene accumulation. Moreover, as a control, pCAROTENE-PTDH3 (Destination vector I expressing all three CDSs from the strong constitutive yeast *TDH3* promoter) was transformed into strains Gen 0.1 (pG1pDI, **Supplementary Fig. 9e**) and IMX672.1 (pMXpDI, **Supplementary Fig. 9f**). Although the same promoter controls expression of the three genes, we observed a small variation in the colony colors. In general, (i) we could not reliably probe the stability and robustness of expression of the assembled pathways after retransformation into yeast cells, because the variation in color (light yellow to dark orange) between colonies may result from a variation in the plasmid copy number. (ii) ATF/BS library- or *TDH3* promoter-containing strains established in the Gen 0.1 background (**Supplementary Fig. 9b,e**) showed better  $\beta$ -carotene accumulation than strains established in strain IMX672.1 (**Supplementary Fig. 9c,f**). Strains Gen 0.1 and IMX672.1 with pDI 1 produced colonies with the most intense color. Two colonies, one from strain Gen 0.1 (G1pDI 1) and one from strain IMX672.1 (MXpDI 1) were selected for quantitative determination of  $\beta$ -carotene content by HPLC (**Supplementary Fig. 9g**). The HPLC data demonstrated that G1pDI 1 cells accumulated slightly more  $\beta$ -carotene ( $0.81 \pm 0.25 \text{ mg g}^{-1} \text{ cdw}$ ) than MXpDI 1 cells ( $0.61 \pm 0.14 \text{ mg g}^{-1} \text{ cdw}$ ); this difference, however, was statistically non-significant based on Student's *t*-test ( $p$ -value > 0.01), and was probably affected by different plasmid copy numbers in the cells.

#### Supplementary Note 4

##### Single-locus integration of a Destination vector I library carrying $\beta$ -carotene pathway genes

The library of  $\beta$ -carotene pathway genes assembled in Destination vector I from Level 2 was integrated into the *lys2.a* locus of strain Gen 0.1 or IMX672.1 (**Supplementary Fig. 10a**). The selection for integration was carried out on SC-Ura/-Leu/-His plates, as the *URA3*, *LEU2*, and *HIS3* genes are encoded on the  $\beta$ -carotene assembly modules. A spectrum of colors ranging from yellow to deep orange was observed for the different colonies (Gen 0.1 background, **Supplementary Fig. 10b**; IMX672.1 background, **Supplementary Fig. 10c**). ATF/BS modules upstream of each CDS of three deep-orange colonies from each plate (IMX672.1 and Gen 0.1) were amplified by PCR and sequenced. We found that weak to strong, medium, and weak or strong ATF/BS regulators, respectively, drive expression in the *BTS1*, *McrTYB* and *McrTI* modules in the Gen 0.1 background (G1intpDI 1, G1intpDI 2, and G1intpDI 3; **Supplementary Fig. 10d**). In the case of strain IMX672.1, *BTS1* expression, recovered from modules of three



independent colonies, was controlled by weak or strong ATF/BS regulators, while all *McrTYB* modules contained strong ATF/BS regulators, and *McrTI* expressing modules contained weak or strong ATF/BSs (MXintpDI 1, MXintpDI 2, and MXintpDI 3; **Supplementary Fig. 10d**). Moreover, as a control, pCAROTENE-PTDH3 (Destination vector I expressing all three  $\beta$ -carotene biosynthesis genes from the strong yeast *TDH3* promoter; see **Online Methods**) was integrated into locus *lys2.a* of strains IMX672.1 and Gen 0.1 to generate, respectively, strains PosG1intpDI (**Supplementary Fig. 10e**) and PosMXintpDI (**Supplementary Fig. 10f**). Almost all yeast colonies were similar in color. In general, (i) ATF/BS library- or *TDH3* promoter-containing strains established in Gen 0.1 (**Supplementary Fig. 10b,e**) showed stronger  $\beta$ -carotene accumulation than strains established in the IMX672.1 background (**Supplementary Fig. 10c,f**). We selected two colonies for quantitative determination of  $\beta$ -carotene content by HPLC, one from strain Gen 0.1 (G1intpDI 1; **Supplementary Fig. 10b**), and one from strain IMX672.1 (MXintpDI 1; **Supplementary Fig. 10c**). The HPLC results (**Supplementary Fig. 10g**) demonstrated a 3.3-fold higher  $\beta$ -carotene level in the background-optimized strain Gen 0.1 ( $0.46 \pm 0.05 \text{ mg g}^{-1} \text{ cdw}$ ) than in the IMX672.1 wild type ( $0.14 \pm 0.05 \text{ mg g}^{-1} \text{ cdw}$ ).

## Supplementary Note 5

### Construction of a library for controllable $\beta$ -ionone production

We established the four-gene pathway required for the production of  $\beta$ -ionone, a downstream product of  $\beta$ -carotene, in the three best  $\beta$ -carotene producers achieved by three approaches. Successful expression of *RiCCD1*, which converts  $\beta$ -carotene to  $\beta$ -ionone, leads to yeast cells that are less intensely colored than  $\beta$ -carotene-producing cells. At Level 0, we cloned the *RiCCD1* CDS into Entry vector X (**Supplementary Table 3**). At Level 1, we assembled the library of ATF/BS units and the *RiCCD1* CDS in Acceptor vector C (**Supplementary Fig. 6**, **Supplementary Fig. 8a**, and **Supplementary Table 4**). The nine diverse ATF/BS control modules upstream of the *RiCCD1* CDS were identified by PCR using primers ATF-for and BS-rev, followed by sequencing (**Supplementary Fig. 8b**). We employed the two yeast strains IMX672.1 and Gen 0.1, and studied the production of  $\beta$ -ionone using methods described in the previous section for analyzing the levels of  $\beta$ -carotene (approach 1, **Supplementary Fig. 11**; approach 2, **Supplementary Fig. 12**; and approach 3, **Supplementary Fig. 13**). Yeast cells expression all for  $\beta$ -ionone CDSs from the *TDH3* promoter were used as positive control. Three weakly colored colonies were chosen to sequence the ATF/BS modules controlling *RiCCD1* expression and one of them was analyzed further by HPLC. We selected the best  $\beta$ -carotene producers in the IMX672.1 and Gen 0.1 backgrounds to introduce the  $\beta$ -ionone library (nine

ATF/BS regulators upstream of *RiCCD1*). Moreover, we generated a 9<sup>4</sup>-member library of  $\beta$ -ionone producing strains to identify a strain producing  $\beta$ -ionone at the highest possible level (described in the next chapter). We observed that approach 3 (**Supplementary Fig. 13**), 1 (**Supplementary Fig. 11**), and 2 (**Supplementary Fig. 12**) resulted in high, medium and low amounts of  $\beta$ -ionone accumulation, respectively. Surprisingly, in all three cases more  $\beta$ -ionone was produced when pathway genes were assembled in wild-type strain IMX672.1 than in the optimized strain Gen 0.1, confirming that the high level of  $\beta$ -carotene produced in strain Gen 0.1 was not efficiently converted to  $\beta$ -ionone. Overall, our top producer (MXintR1pDI 1; **Supplementary Fig. 13g**) yielded 3.3-fold more  $\beta$ -ionone than the previously reported strain *RiCCD1*<sup>1</sup> ( $0.042 \pm 0.018$  mg g<sup>-1</sup> cdw in our experiments; **Supplementary Table 6**).

### Supplementary Note 6

#### Construction of a library for $\beta$ -ionone pathway genes by multi-locus integration

The library of ATF/BS-*RiCCD1* CDS modules was integrated into the genomic *ura3-52* locus (**Supplementary Fig. 11a**) of two strains producing high levels of  $\beta$ -carotene, namely CB1 (Gen 0.1 background, **Supplementary Fig. 11b**) and CC1 (IMX672.1 background, **Supplementary Fig. 11c**). The selection for successful assemblies was carried out by plating yeast cells on SC-URA-LEU/-HIS/-TRP/-LYS medium, as the *URA3*, *LEU2*, *HIS3* and *TRP1* genes are encoded by the  $\beta$ -carotene and  $\beta$ -ionone assembly modules and *LYS2* is encoded on the Cas9/gRNA expression plasmid (see **Online Methods**). The enzyme *RiCCD1* converts  $\beta$ -carotene to  $\beta$ -ionone whereby yeast colonies lose their deep-orange color. DNA of three light-yellow colonies from each plate (IMX672.1 and Gen 0.1) was used for PCR amplification of ATF/BS modules upstream of the *RiCCD1* CDS, and PCR-amplified fragments were sequenced (**Supplementary Fig. 11d**). We found that a weak to medium ATF/BS regulator driving *RiCCD1* expression results in highest  $\beta$ -ionone accumulation in the Gen 0.1 background (RB1, RB2, and RB3; **Supplementary Fig. 11b,d**), while a medium to strong ATF/BS regulator upstream of *RiCCD1* results in high  $\beta$ -ionone accumulation in the IMX672.1 yeast (RC1, RC2, and RC3; **Supplementary Fig. 9c,d**). Moreover, three modules expressing *RiCCD1* from the strong yeast *TDH3* promoter was integrated in the same locus of Gen 0.1 (PRB, **Supplementary Fig. 11e**) and IMX672.1 (PRC, **Supplementary Fig. 11f**). Almost all colonies were similar in color. We selected two colonies for quantitative determination of  $\beta$ -ionone, *i.e.* colony RB1 from strain Gen 0.1 (**Supplementary Fig. 11b**), and colony RC1 from strain IMX672.1 (**Supplementary Fig. 9c**). As strains producing high levels of  $\beta$ -carotene (CB1 and CC1) were chosen for genomic integration of the *RiCCD1* CDS (see above), comparing the  $\beta$ -carotene levels in the two host

strains with those of the  $\beta$ -ionone-producing strains indicates the extent of  $\beta$ -carotene usage. The HPLC data (**Supplementary Fig. 11g**) demonstrate a 20- and 14-fold lower  $\beta$ -carotene level in  $\beta$ -ionone expressing strains, RB1 and RC1, than in the backbone strains, CB1 and CC1, respectively, allowing to produce  $0.060 \pm 0.008 \text{ mg g}^{-1} \text{ cdw}$  and  $0.062 \pm 0.0001 \text{ mg g}^{-1} \text{ cdw}$  of  $\beta$ -ionone in RB1 and RC1, respectively. The highest  $\beta$ -ionone production of  $0.22 \pm 0.06 \text{ mg g}^{-1} \text{ cdw}$  was previously reported for strain RiCCD1<sup>1</sup>, while it produced  $0.042 \pm 0.018 \text{ mg g}^{-1} \text{ cdw}$  of  $\beta$ -ionone in our experiments (**Supplementary Table 6**). This represents a 1.5-fold improvement of strain RC1 over RiCCD1.

### Supplementary Note 7

#### Construction of a Destination vector library for controllable $\beta$ -ionone production

A library of ATF/BSs-*RiCCD1* modules was assembled in vector pDI 1 (Destination vector I-BTS1-McrtI-McrtYB leading to high  $\beta$ -carotene production) using TAR to generate an episomal plasmid library for  $\beta$ -ionone production (**Supplementary Fig. 12a**), using strain Gen 0.1 as the host. The generated library was called RpDI library. The selection for clones with correctly assembled pathways was carried out by plating the yeast cells on medium lacking the auxotrophic markers (SC-Ura/-Leu/-His/-Trp); the corresponding genes are encoded by the assembly modules. A variety of colors ranging from light yellow to orange was observed in the different colonies (**Supplementary Fig. 12b**). As the color of the colonies may result from the presence of different numbers of plasmids within a given yeast cell, plasmids were recovered and transformed into *E. coli*. The structure of three plasmids was confirmed by sequencing; the plasmids were then retransformed into IMX672.1 and Gen 0.1 to confirm functional expression of the pathway. Our results showed that medium to strong ATF/BS regulators resulted in colonies with reduced color formation in both strains (MXR1pDI 1 and G1R1pDI 1; **Supplementary Fig. 12c**). Moreover, as a control, pIONONE-PTDH3 (Destination vector I expressing all four  $\beta$ -ionone biosynthesis genes from the strong yeast *TDH3* promoter; see **Online Methods**) was transformed into strain Gen 0.1 (PosG1pDI; **Supplementary Fig. 12d**). Despite the fact that the same promoter controlled the expression of all four genes, we observed some variation in the colony colors likely due to different plasmid copy numbers. We analyzed the  $\beta$ -ionone content in the G1R1pDI 1 and MXR1pDI 1 strains using HPLC (**Supplementary Fig. 12g**). The data demonstrate that the amount of  $\beta$ -carotene in strain MXpDI 1 (parental strain with a CEN.PK background) was decreased 20-fold to produce  $0.039 \pm 0.05 \text{ mg g}^{-1} \text{ cdw}$  of  $\beta$ -ionone in strain MXR1pDI 1, while the  $\beta$ -carotene amount in strain G1pDI 1 (parental strain with the optimized biochemical background) was decreased 16-fold to produce  $0.036 \pm 0.002$

mg g<sup>-1</sup> cdw β-ionone in strain G1R1pDI 1, demonstrating that the high amount of β-carotene cannot be efficiently converted to the downstream product.

### Supplementary Note 8

#### Single-locus integration of a Destination vector I library carrying β-ionone pathway genes

In the next step, the RpDI library containing the nine ATF/BS regulators to control *RiCCD1* expression was integrated into the *lys2.a* loci of strains IMX672.1 and Gen 0.1 (**Supplementary Fig. 12a**). The selection of successful assemblies was carried out by plating the yeast cells on SC-Ura/-Leu/-His/-Trp medium, as the *URA3*, *LEU2*, *HIS3* and *TRP1* genes are encoded by the assembly modules. A variety of colony colors ranging from light yellow to orange was observed (IMX672.1, **Supplementary Fig. 12b**; and Gen 0.1, **Supplementary Fig. 12c**). We checked the sequences of ATF/BS regulators leading to high β-ionone levels. Our result showed that *RiCCD1* modules need less strong ATF/BS regulators to produce high β-ionone levels in the presence of high precursor levels in the Gen 0.1 background in comparison to IMX672.1. Moreover, as a control, pIONONE-PTDH3 was integrated into the same chromosomal locus of the strains Gen 0.1 (PosG1intRpDI, **Fig. Supplementary 12e**) and IMX672.1 (PosMXintRpDI, **Fig. Supplementary 12f**). Almost all colonies grown on solid medium were of similar color. In general, ATF/BS library- or *TDH3* promoter-containing strains established in strain Gen 0.1 (**Supplementary Fig. 12b,e**) showed better β-carotene consumption (and therefore accumulation of β-ionone) than strains established in strain IMX672.1 (**Supplementary Fig. 12c,f**). We selected two colonies for quantitative determination of β-ionone content by HPLC (**Supplementary Fig. 12g**), one from strain Gen 0.1 (G1intR1pDI 1) and one from strain IMX672.1 (MXintR1pDI 1). The HPLC data demonstrate that the amount of β-carotene in strain MXintpDI 1 (wild-type background) was decreased 7-fold to produce 0.14 ± 0.02 mg g<sup>-1</sup> cdw β-ionone in MXintR1pDI 1, while the β-carotene amount in strain G1intpDI 1 (optimized background) was decreased 15-fold to produce 0.12 ± 0.06 mg g<sup>-1</sup> cdw β-ionone in G1intR1pDI 1. This result is in accordance with our previous observation that a high level of precursor (β-carotene) does not necessarily result in high amounts of the downstream metabolite (β-ionone). In conclusion, we generated a β-ionone producing yeast strain, named MXintR1pDI 1, which accumulates 3.3-fold more β-ionone than the previously reported strain RiCCD1<sup>1</sup>.

### References

1. Beekwilder, J. *et al.* Polycistronic expression of a beta-carotene biosynthetic pathway in *Saccharomyces cerevisiae* coupled to beta-ionone production. *J Biotechnol* **192 Pt B**, 383-392 (2014).

## Supplementary Protocols

### Combinatorial assembly of plant-derived ATFs and BSs in Entry Vector X

1. Expression plasmids<sup>1</sup> containing the coding sequences (CDSs) of plant-derived artificial transcription factors (ATFs) are mixed in equimolar ratio and DNA fragments harboring the IPTG-inducible *GAL1* promoter, the CDSs of plant derived ATFs and the *CYC1* terminator (*Pro<sub>mGAL1-LacI</sub>-ATF-Ter<sub>CYC1</sub>*) are amplified by multiplex-PCR using primers ATF\_for and ATF\_rev.
  - Primer ATF\_for overlaps with the *X0* region.
  - Primer ATF\_rev overlaps with primer BS\_for.
2. Reporter plasmids<sup>1</sup> containing the binding sites (BSs) of ATFs are mixed in equimolar ratios and DNA fragments harboring one, two or four copies of the binding sites (*Pro<sub>CYC1min-BS</sub>*) are obtained by multiplex-PCR using primers BS\_for and BS\_rev.
  - Primer BS-rev overlaps with the *Z0* region, representing the last 30 bp of the minimal *CYC1* promoter.
3. Entry vector X is digested at the multiple cloning site (*MCS*).
4. Multiplex-PCR-amplified ATFs, their corresponding BS fragments, and linearized Entry vector X are mixed at a concentration ratio of 2:10:1 to perform NEBuilder HiFi DNA assembly in a single reaction tube.
5. The cloning reaction is transformed into *E. coli* and plated onto LB agar media containing Kanamycin (50 µg/ml).
6. Colony-PCR is performed using primers EXSEQ-for/EXSEQ-rev and primers PROSEQ-for/PROSEQ-rev to identify the respective combinations of ATFs and BSs, respectively.
7. Desired colonies are inoculated into LB liquid medium containing Kanamycin (50 µg/ml) for sequencing to identify the respective combinations of ATFs and BSs.
  - Sequences *X0* and *Z0* are explained in the footnote to **Supplementary Table 2**.

### Design of primers for the assembly of CDS units in Entry vector X

1. Forward primer for PCR amplification of CDSs:
  - The upstream (left) homology region (HR) of the linearized vector (called *Z0*) defines the upstream (left) HR of the primer.
  - A rare restriction enzyme (RE) cleavage site is introduced upstream the gene's translation start codon that is compatible to the next level vectors (Destination vectors

I/II: *Asi*/SI; Acceptor vectors A/E: *I-CeuI*, B/F: *I-SceI*, C/G: *PI-PspI*, and E/F: *PI-SceI/AscI*). For optional possibilities see end of this section.

2. Reverse primer for PCR amplification of CDSs:
  - The downstream (right) HR of the primer is defined based on the upstream region of the chosen yeast terminator.
3. Forward primer for PCR amplification of the terminator:
  - The primer is defined based on the upstream region of the chosen yeast terminator.
4. Reverse primer for PCR amplification of the terminator:
  - The same rare RE recognition site as above is introduced downstream of the terminator. For optional possibilities see end of this section.
5. Forward primer for PCR amplification of the promoter of the *E. coli* selection marker:
  - The upstream (left) HR of the primer overlaps with the yeast terminator.
6. Reverse primer for PCR amplification of the promoter of the *E. coli* selection marker:
  - The downstream (right) HR of the primer overlaps with the *Y0* region of the vector and introduces Z (*Z1*, *Z2*, *Z3*, *Z4*, or *Z5*) sequences, representing the first 30 bp of the CDSs of the *E. coli* selection markers encoded by the next-level vectors.

Sequences *Y0* and *Z0* - *Z5* are explained in the footnote to **Supplementary Table 2**.

Optional: Introducing the RE sites in the primer sequences allows replacement of the CDS(s) and yeast terminator(s) with other CDS(s) and terminator(s) at a later stage when alternative COMPASS vectors are needed.

### Combinatorial assembly of ATF/BS units and genes in Level 1 vectors

1. Nine Entry vectors X-ATF/BS containing the CDSs of plant-derived ATFs and BSs are mixed in equimolar ratio. The ATF/BS units (*Pro<sub>mGAL1-LacI</sub>-ATF-Ter<sub>CYC1</sub>-Pro<sub>CYC1min-BS</sub>*) are PCR-amplified using primers X0\_for/Z0\_rev.
  - Primer X0\_for overlaps with the *X0* region in the vectors.
  - Primer Z0\_rev overlaps with the last 30 bp of the *CYC1* promoter.
2. Five Entry vectors X-CDS are mixed in equimolar ratio. The CDS units (*CDS-Ter<sub>yeast</sub>-Pro<sub>E.coli</sub>*) are PCR-amplified using primers Z0\_for, Z1\_rev, Z2\_rev, Z3\_rev, Z4\_rev and Z5\_rev at a concentration ratio of 5:1:1:1:1:1.
  - Primer Z0\_for, representing the last 30 bp of the *CYC1* promoter, overlaps with primer Z0\_rev.
  - Primers Z1\_rev, Z2\_rev, Z3\_rev, Z4\_rev, and Z5\_rev overlap with the Z1 (Destination vector I), Z2 (Acceptor vectors A), Z3 (Acceptor vectors B), Z4 (Acceptor vectors C),

and Z5 (Acceptor vectors D), representing the first 30 bp of the CDSs of the spectinomycin (*Spect<sup>R</sup>*), ampicillin (*Amp<sup>R</sup>*), chloramphenicol (*Cm<sup>R</sup>*), triclosan (*TCS<sup>R</sup>*), or gentamicin (*Gen<sup>R</sup>*) resistance gene, respectively.

3. The *Sall*/*EcoRI*-digested Destination vector I and *FseI*/*AscI*-digested Acceptor vectors A, B, C, and D are mixed in equimolar ratio.
  4. Multiplex PCR-amplified (nine) ATF/BS units, multiplex PCR-amplified (five) CDS units, and (five) linearized vectors are mixed at a concentration ratio of 2:2:1 to perform NEBuilder HiFi DNA assembly in a single reaction tube.
  5. The reaction cocktail is transformed into *E. coli*.
  6. The *E. coli* cells are plated onto LB agar media containing either spectinomycin (50 µg/ml), ampicillin (50 µg/ml), chloramphenicol (25 µg/ml), triclosan (14.5 µg/ml), or gentamicin (50 µg/ml) to generate Destination vectors I-(ATF/BS-)CDS1, and Acceptor vectors A-(ATF/BS-)CDS2, B-(ATF/BS-)CDS3, C-(ATF/BS-)CDS4, and D-(ATF/BS-)CDS5.
  7. Colony-PCR is performed using primers ATF-for and BS-rev.
  8. Desired colonies are inoculated into LB liquid medium containing appropriate antibiotic (see step 6) for sequencing to identify the respective combinations of ATFs and BSs.
- Sequences X0 and Z0 - Z5 are explained in the footnote to **Supplementary Table 2**.

### **Design of primers for combinatorial assembly of gene units in Level 1 vectors**

1. Forward primer for PCR amplification of CDSs:  
The upstream (left) HR of the linearized vector (called X0) defines the upstream (left) HR of the primer.  
A rare RE cleavage site is introduced upstream the gene's translation start codon that is compatible to the vectors (Destination vectors I/II: *AsiSI*, Acceptor vectors A/E: *I-CeuI*, B/F: *I-Scel*, C/G: *PI-PspI*, and E/F: *PI-Scel/AscI*). For optional possibilities see end of this section.
2. Reverse primer for PCR amplification of CDSs:  
The downstream (right) HR of the primer is defined based on the upstream region of the chosen yeast terminator.
3. Forward primer for PCR amplification of the terminator:  
The primer is defined based on the upstream region of the chosen yeast terminator.
4. Reverse primer for PCR amplification of the terminator:  
The same rare RE recognition site as above is introduced downstream of the terminator. For optional possibilities see end of this section.
5. Forward primer for PCR amplification of the promoter of the *E. coli* selection marker:  
The upstream HR of the primer overlaps with the yeast terminator.



6. Reverse primer for PCR amplification of the promoter of the *E. coli* selection marker:

The downstream (right) HR of the primer overlaps with Z1 (Destination vector I), Z2 (Acceptor vectors A), Z3 (Acceptor vectors B), Z4 (Acceptor vectors C), and Z5 (Acceptor vectors D), representing the first 30 bp of the CDSs of *Spect<sup>R</sup>*, *Amp<sup>R</sup>*, *Cm<sup>R</sup>*, *TCS<sup>R</sup>*, or *Gen<sup>R</sup>*, respectively.

Sequences X0 and Z0 - Z5 are explained in the footnote to **Supplementary Table 2**.

Optional: Introducing the RE sites in the primer sequences allows replacement of the CDS(s) and yeast terminator(s) with other CDS(s) and terminator(s) at a later stage when alternative COMPASS vectors are needed.

### **Combinatorial assembly of pathways in Destination vector I at Level 2**

1. Destination vectors I-(ATF/BS-)CDS1 are mixed at equimolar concentration and digested with *I-CeuI* or/and *SbfI*.
2. The library of Acceptor vectors A-(ATF/BS-)CDS2 is mixed at equimolar concentration and used for multiplex PCR amplification of ATF/BS-CDS2 modules (*Pro<sub>HIS3</sub>-Pro<sub>mGAL1</sub>-ATF/BS-CDS2-Ter<sub>yeast</sub>-Pro<sub>AmpR</sub>-Amp<sup>R</sup>-Ter<sub>AmpR</sub>*) using primers X0\_for and Y1\_rev.
3. Using TAR, the PCR products are assembled in the library of Destination vectors-(ATF/BS-)CDS1 to generate a library of Destination vectors I-(ATF/BS-)CDS1-(ATF/BS-)CDS2.
4. Yeast cells are plated on SC-Ura/-His medium.
5. After 4 d, cells are scraped from the plates and the plasmid library is extracted.
6. The recovered plasmid library is transformed into *E. coli*, and cells are grown on LB agar plates containing ampicillin (50 µg/ml).
7. *E. coli* cells are scraped from the plates and the plasmid library is extracted.
8. The library of Acceptor vectors B-(ATF/BS-)CDS3 is used for multiplex PCR amplification of ATF/BS-CDS3 modules (*Pro<sub>LEU2</sub>-Pro<sub>mGAL1</sub>-ATF/BS-CDS3-Ter<sub>yeast</sub>-Pro<sub>CmR</sub>-Cm<sup>R</sup>-Ter<sub>CmR</sub>*) using primers X0\_for and Y2\_rev.
9. Using TAR, the PCR products are assembled in the library of Destination vectors-(ATF/BS-)CDS1-(ATF/BS-)CDS2 digested with *I-SceI* or/and *FseI* to generate a library of Destination vectors I-(ATF/BS-)CDS1-(ATF/BS-)CDS2-(ATF/BS-)CDS3.
10. Yeast cells are plated on SC-Ura/-His/-Leu medium.
11. After 4 d, cells are scraped from the plates and the plasmid library is extracted.
12. The recovered plasmid library is transformed into *E. coli*, and cells are grown on LB agar plates containing chloramphenicol (25 µg/ml).
13. *E. coli* cells are scraped from the plates and the plasmid library is extracted.

14. The library of Acceptor vectors C-(ATF/BS-)CDS4 is used for multiplex PCR amplification of ATF/BS-CDS4 modules (*Pro<sub>TRP1</sub>-Pro<sub>mGAL1</sub>-ATF/BS-CDS4-Ter<sub>yeast</sub>-Pro<sub>TCSR</sub>-TCS<sup>R</sup>-Ter<sub>TCSR</sub>*) using primers X0\_for and Y3\_rev.
15. Using TAR, the PCR products are assembled in the library of Destination vectors-(ATF/BS-)CDS1-(ATF/BS-)CDS2-(ATF/BS-)CDS3 digested with *Pi-Pspl* or/and *Sfill* to generate a library of Destination vectors I-(ATF/BS-)CDS1-(ATF/BS-)CDS2-(ATF/BS-)CDS3-(ATF/BS-)CDS4.
16. Yeast cells were plated on SC-Ura/-His/-Leu/-Trp medium.
17. After 4 d, cells are scraped from the plates and the plasmid library is extracted.
18. The recovered plasmid library is transformed into *E. coli*, and cells are grown on LB agar plates containing triclosan (14.5 µg/ml).
19. *E. coli* cells are scraped from the plates and the plasmid library is extracted.
20. The library of Acceptor vectors D-(ATF/BS-)CDS5 is used for multiplex PCR amplification of ATF/BS-CDS5 modules (*Pro<sub>LYS2</sub>-Pro<sub>mGAL1</sub>-ATF/BS-CDS5-Ter<sub>yeast</sub>-Pro<sub>GenR</sub>-Gen<sup>R</sup>-Ter<sub>GenR</sub>*) using primers X0\_for and Y4\_rev.
21. Using TAR, the PCR products are assembled in the library of Destination vectors-(ATF/BS-)CDS1-(ATF/BS-)CDS2-(ATF/BS-)CDS3-(ATF/BS-)CDS4 digested with *AscI* to generate a library of Destination vectors I-(ATF/BS-)CDS1-(ATF/BS-)CDS2-(ATF/BS-)CDS3-(ATF/BS-)CDS4-(ATF/BS-)CDS5.
22. Yeast cells are plated on SC-Ura/-His/-Leu/-Trp/-Lys medium.
23. After 4 d, cells are scraped from the plates and plasmid library is extracted.
24. The recovered plasmid library is transformed into *E. coli*, and cells are grown on LB agar plates containing gentamicin (50 µg/ml).
25. *E. coli* cells are scraped from the plates and the plasmid library is extracted.
  - Sequence X0 is explained in the footnote to **Supplementary Table 2**.
  - Sequences Y1, Y2, Y3, and Y4 represent to the last 30 bp of the terminator sequences of the *Amp<sup>R</sup>*, *Cm<sup>R</sup>*, *TCS<sup>R</sup>*, or *Gen<sup>R</sup>* gene, respectively.

## References

1. Naseri, G., Balazadeh, S., Machens, F., Kamranfar, I., Messerschmidt, K. & Mueller-Roeber, B. Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth Biol* **6**, 1742-1756 (2017).

**6 CaPRedit: Genome editing using CRISPR-Cas9 and plant-derived transcriptional regulators for the redirection of flux through the FPP branch-point in yeast**

## ABSTRACT

Technologies developed over the past decade have made *Saccharomyces cerevisiae* a promising platform for production of different natural products. We developed CRISPR/Ca9- and plant derived regulator-mediated genome editing approach (CaPRedit) to greatly accelerate strain modification and to facilitate very low to very high expression of key enzymes using inducible regulators. CaPRedit can be implemented to enhance the production of yeast endogenous or heterologous metabolites in the yeast *S. cerevisiae*. The CaPRedit system aims to facilitate modification of multiple targets within a complex metabolic pathway through providing new tools for increased expression of genes encoding rate-limiting enzymes, decreased expression of essential genes, and removed expression of competing pathways. This approach is based on CRISPR/Cas9-mediated one-step double-strand breaks to integrate modules containing IPTG-inducible plant-derived artificial transcription factor and promoter pair(s) in a desired locus or loci. Here, we used CaPRedit to redirect the yeast endogenous metabolic flux toward production of farnesyl diphosphate (FPP), a central precursor of nearly all yeast isoprenoid products, by overexpression of the enzymes lead to produce FPP from glutamate. We found significantly higher  $\beta$ -carotene accumulation in the CaPRedit-mediated modified strain than in the wild type (WT) strain. More specifically, CaPRedit\_FPP 1.0 strain was generated, in which three genes involved in FPP synthesis, *tHMG1*, *ERG20*, and *GDH2*, were inducibly overexpressed under the control of strong plant-derived ATFPs. The  $\beta$ -carotene accumulated in CaPRedit\_FPP 1.0 strain to a level 1.3-fold higher than the previously reported optimized strain that carries the same overexpressed genes (as well as additional genetic modifications to redirect yeast endogenous metabolism toward FPP production). Furthermore, the genetic modifications implemented in CaPRedit\_FPP 1.0 strain resulted in only a very small growth defect (growth rate relative to the WT is  $\sim -0.03$ ).

## INTRODUCTION

The well-developed genetics and tools for genetic manipulation, solid knowledge of endogenous metabolism, and robustness in large-scale fermentation make traditional baker's yeast, *Saccharomyces cerevisiae*, a suitable cell factory for metabolic engineering (1). As such, it has been broadly used to produce diverse products, e.g. alcohols (2), hydrocarbons (3), and proteins (4). In addition to introducing heterologous (non-native) metabolic pathways, increasing the concentration of endogenous (native) precursors and cofactors may enhance production of desired metabolites (5).

The production of biotechnological products relies on metabolic and genetic engineering tools. Plasmid-based systems are the current standard for overexpression of genes, however their use raises major practical concerns about genetic stability and operational repeatability (6). Hence, expression from donors integrated into the yeast genome is generally preferred (7). However, complex genome engineering projects are often too difficult to be practical because the efficiency of simultaneous integration of multiple DNA fragments is often too low. Here, we used the CRISPR/Cas9 system (8) for genome modification in yeast. The CRISPR/Cas9 strategy is based on the type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR associated proteins (Cas) system and allows simultaneous modification of multiple *S. cerevisiae* genes in one step (8). Not only is this a rapid integration strategy, but it is also efficient enough to exclude a subsequent selection step.

Constitutive promoters are often used for the overexpression or downregulation of enzyme-encoding genes (5,9), but constitutive alteration of enzyme levels may create a metabolic burden on the cell (10). Therefore, inducible promoters are often preferred. As the number of inducible promoters for use in yeast is currently limited (11,12), programmable (and, therefore, artificial) transcription factors (ATFs) have recently been developed (13-15). Although ATFs of different formats, including those build around transcription activator-like effector (TALE)- and CRISPR/dead CRISPR-associated protein 9 (dCas9) scaffolds, have been established in recent years, little attention has so far been paid towards implementing them in metabolic engineering projects. This may be due to the fact that generating TALE-based ATFs is laborious and transcriptional activation by CRISPR/dCas9-derived ATFs is generally lower than that of TALE-derived ATFs (17). We recently demonstrated that plant-derived ATFs are a promising alternative to TALE- and CRISPR/dCas9-derived ATFs due to their relatively small size, ease of construction, and a wider range of transcriptional activities that can be achieved for synthetic biology applications (16). Our CaPRedit approach reported here combines CRISPR/Cas9-mediated one-step genomic integration of multiple modules consisting of (i) plant-derived ATFs and biosynthetic pathway genes, whereby the ATFs are utilized for the inducible overexpression of the pathway genes, and (ii) synthetic promoters containing plant TF binding sites for the downregulation of endogenous genes, whereby the native promoters are replaced with synthetic promoters. Moreover, we used a high-copy-number plasmid (with a  $2\mu$  origin) carrying iCas9, a variant of Cas9 which has a higher activity than wild-type nuclease.

The mevalonate (MVA) biosynthesis pathway converts acetyl-coenzyme A (acetyl-CoA) to farnesyl pyrophosphate (FPP), representing a key branch point for the biosynthesis of all isoprenoids. MVA is typically found in many eukaryotes, including plants and yeast, archaea, and some eubacteria. One subgroup of the isoprenoids are the carotenoids which comprise over

700 different chemical structures (20).  $\beta$ -Carotene is non-native to *S. cerevisiae* but is a carotenoid found in plants. It is known as a pro-vitamin A carotenoid with anti-oxidant activity and is furthermore used to color foods. To achieve high-level production of such a non-native isoprenoid in *S. cerevisiae*, the flux of FPP must be diverted toward the heterologous metabolic reactions.

CaPRedit can be employed to increase the product output of any engineered biosynthetic pathway in yeast by increasing the concentrations of endogenous precursors or by expressing exogenous enzymes. Here, as a proof of concept, we used CaPRedit to improve the production of FPP. Previously engineered yeast strains redirected the flux of FPP toward  $\beta$ -carotene (5,18) and other secondary metabolites (12,19) by redirecting yeast endogenous metabolites toward FPP and subsequently FPP toward the desired secondary metabolites. The over-accumulation of FPP, and its redirection toward carotenoid production, was achieved in strain SCIGS22a (with CEN.PK background) (5) via three groups of modifications including the overexpression of enzymes to increase production of FPP, and the deletion and downregulation of enzymes that consume FPP for alternative pathways. More specifically, (i) enzymes of the MVA biosynthesis pathway were overexpressed, including the following: variants of the native yeast ERG20 enzyme (FPP synthase) that converts isopentenyl diphosphate (IPP) to geranyl diphosphate (GPP) and FPP; NAD-dependent glutamate dehydrogenase (GDH2) which consumes NADH during the anabolic process and changes the yeast cofactor balance; and catalytic domain of 3-hydroxy-3-methyl-glutaryl-CoA reductase (tHMG1). Previous studies showed that overexpression of tHMG1 results in an increased production of isoprenoids (21), while another report showed that the full-length version was more effective toward the production of prenol alcohols (22). Hence, we used tHMG1 in this study; (ii) deletion of lipid phosphate phosphatase 1 (*LPP1*) and diacylglycerol diphosphate phosphatase 1 (*DPP1*) genes to minimize farnesol formation from FPP, and deletion of NADP<sup>+</sup>-glutamate dehydrogenase 1 (*GDH1*). The conversion of HMG-CoA to mevalonate requires NADPH, thus the deletion of *GDH1* was expected to redirect carbon flux through the mevalonate pathway by increasing the pool of available NADPH for HMGR (23); (iii) downregulation of squalene synthase *ERG9*, which converts FPP to sterols and farnesols which are essential metabolites in yeast but undesirable products when optimizing FPP production is the goal to increase  $\beta$ -carotene production.

Here, we used CaPRedit to modify yeast strain IMX672.1 (CEN.PK background) to generate a new strain, dubbed CaPRedit\_FPP 1.0, in which three enzymes, *i.e.* ERG20, GDH2, and tHMG1 are inducibly overexpressed by strong plant-derived ATFs. After induction of the ATFs in CaPRedit\_FPP 1.0 engineered to inducibly express three-gene  $\beta$ -carotene pathway,  $\beta$ -carotene

production improved by 4.3- and 1.3-fold compared to the parental strain (IMX672.1: IMX672 background; **Table 1** and **2**, **Supplementary Methods**) and the strain previously optimized for FPP production, Gen 0.2 (SCIGS22a (18) background; **Table 1** and **2**, **Supplementary Methods**). The  $\beta$ -carotene production can be increased even further if, in addition to the overexpression of genes of the MVA pathway, expression of *ERG9* is downregulated and the genes *LPP1*, *DPP1*, and *GDH1* are inactivated or deleted. Our achievements can pave the way for the application of our collection of plant-derived ATFs for either improving the production of endogenous precursors or expression of heterologous pathways in yeast.

## MATERIALS AND METHODS

### General

Plasmids were constructed by either NEBuilder HiFi DNA assembly (New England Biolabs, Frankfurt am Main, Germany) or SLiCE (24,25). Primer and plasmid sequences are given in **Supplementary Table S1** and **Supplementary Table S2**, respectively; plasmid maps are shown in **Supplementary Figure S1**. PCR amplifications were done using Phusion (LifeTechnologies, Darmstadt, Germany) or PrimeSTAR GXL DNA Polymerase (Takara Bio, Saint-Germain-en-Laye, France). Amplified DNA parts were gel-purified prior to further use. Oligonucleotides were purchased from Eurofins Genomics (Ebersberg, Germany). All constructs were confirmed by sequencing (LGC Genomics, Berlin, Germany).

### Bacterial and yeast strains

Plasmids were transformed into *Escherichia coli* *NEB 5 $\alpha$*  or *NEB 10 $\beta$*  cells (New England Biolabs). Strains were grown in Luria-Bertani medium with appropriate *E. coli* selection markers at 37°C.

*S. cerevisiae* strains YPH500 (ATCC: 76626; www.atcc.org), IMX672 (Euroscarf, #Y40595; www.euroscarf.de), and Gen 0.2 (modified SCIGS22a (5,18), see **Supplementary Methods**) were used. Information on yeast strains constructed in this work is presented in **Table 1**. Generating competent yeast cells and genetic transformation of plasmids or linearized DNA fragments were done using the LiAc/SS carrier DNA/PEG method (26), or the Frozen-EZ Yeast Transformation II Kit (Zymo Research, Freiburg, Germany). Strains were grown at 30°C in yeast extract peptone dextrose adenine (YPDA)-rich medium, or in appropriate synthetic complete (SC) media lacking one or more amino acids to allow selection for transformed cells. Verification of transformation was done using colony-PCR, followed by sequencing of PCR-amplified DNA fragments. Either Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research Corporation) or the

method described by Noskov *et al.* (27) were used to recover plasmids from positive yeast clones.

**Table 1.** List of *S. cerevisiae* strains used in this study.

| Yeast strain              | Relevant genome   | Origin     |
|---------------------------|---|------------|
| YPH500                    | <i>MAT a, URA3-52, lys2-801, ade2-101, trp1D63, his3D200, leu2D1</i>  | ATCC       |
| IMX672                    | <i>MAT a, MAL2-8<sup>c</sup>, SUC2, his3D1, leu2-3_112, ura3-52, trp1-289, can1D::Cas9-natNT2</i>   | Euroscarf  |
| IMX672.1                  | IMX672 + <i>lys2a</i>   | This study |
| Gen 0.1                   | SCIGS22a(5) + <i>lys2a, trp1a, his3a, leu2a</i>   | This study |
| Gen 0.2                   | Gen 0.1 + <i>can1.w::cas9</i>   | This study |
| CapRedit 1.0_FPP          | IMX672 + ADE2:: <i>P<sub>TEF1</sub>-LacI-T<sub>ADH1</sub>-P<sub>GAL1-lacO</sub>-NLS-JUB1-GAL4AD-T<sub>CYC1</sub>/2X-GDH2-T<sub>GDH2</sub>, ura3-52:: P<sub>TEF1</sub>-LacI-T<sub>ADH1</sub>-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-ATAF1-T<sub>CYC1</sub>/2X-ERG20-T<sub>ERG20</sub>, his3D1::P<sub>TEF1</sub>-LacI-T<sub>TEF1</sub>-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-ANAC102-T<sub>CYC1</sub>/4X-tHMG1-T<sub>tHMG1</sub></i>                   | This study |
| Red_IMX672.1              | IMX672.1 + <i>ura3-52:: Leu2-P<sub>TEF2</sub>-DsRed-T<sub>TEF2</sub></i>  | This study |
| Green_IMX672.1            | IMX672.1 + <i>ura3-52:: Leu2-P<sub>TEF2</sub>-AcGFP1-T<sub>TEF2</sub></i>   | This study |
| Red_Gen 0.2               | Gen 0.2 + <i>ura3-52:: Leu2-P<sub>TEF2</sub>-DsRed-T<sub>TEF2</sub></i>   | This study |
| Red_CaPRedit_FPP 1.0      | CaPRedit 1.0_FPP + <i>ura3-52::Leu2-P<sub>TEF2</sub>-DsRed-T<sub>TEF2</sub></i>   | This study |
| IMX672.1_carotene         | IMX672.1 + XII-5:: <i>P<sub>TEF1</sub>-LacI-T<sub>ADH1</sub>-URA3-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-GRF7-T<sub>CYC1</sub>/4X-McrtI-T<sub>FBA1</sub>, XI-2:: P<sub>TEF1</sub>-LacI-T<sub>ADH1</sub>-HIS3-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-GRF7-T<sub>CYC1</sub>/4X-BTS1-T<sub>TDH3</sub>, X-3::P<sub>TEF1</sub>-LacI-T<sub>TEF1</sub>-LEU2-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-RAV1-T<sub>CYC1</sub>/4X-McrtYB-T<sub>SynTer3</sub></i>         | This study |
| CaPRedit_FPP 1.0_carotene | CaPRedit_FPP 1.0 + XII-5:: <i>P<sub>TEF1</sub>-LacI-T<sub>ADH1</sub>-URA3-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-GRF7-T<sub>CYC1</sub>/4X-McrtI-T<sub>FBA1</sub>, XI-2:: P<sub>TEF1</sub>-LacI-T<sub>ADH1</sub>-HIS3-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-GRF7-T<sub>CYC1</sub>/4X-BTS1-T<sub>TDH3</sub>, X-3::P<sub>TEF1</sub>-LacI-T<sub>TEF1</sub>-LEU2-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-RAV1-T<sub>CYC1</sub>/4X-McrtYB-T<sub>SynTer3</sub></i> | This study |
| Gen 0.2_carotene          | Gen 0.2 + XII-5:: <i>P<sub>TEF1</sub>-LacI-T<sub>ADH1</sub>-URA3-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-GRF7-T<sub>CYC1</sub>/4X -McrtI-T<sub>FBA1</sub>, XI-2:: P<sub>TEF1</sub>-LacI-T<sub>ADH1</sub>-HIS3-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-GRF7-T<sub>CYC1</sub>/4X-BTS1-T<sub>TDH3</sub>, X-3::P<sub>TEF1</sub>-LacI-T<sub>TEF1</sub>-LEU2-P<sub>GAL1-lacO</sub>-NLS-GAL4AD-RAV1-T<sub>CYC1</sub>/4X-McrtYB-T<sub>SynTer3</sub></i>         | This study |

### Construction of plasmids and donors for overexpression

Coding sequences of JUB1-, ANAC102-, and ATAF1-derived ATFs were obtained by PCR using appropriate plasmids (16) as templates and the respective forward (ATF-for) and reverse (ATF-



rev) primers. The corresponding binding sites (*JUB1 2X*, *ANAC102 4X*, and *ATAF1 2X*) fused to the yeast minimal *CYC1* promoter were obtained by PCR using appropriate reporter plasmids (16) as templates and the respective forward (BS-for) and reverse (BS-rev) primers. Both fragments were inserted into the Entry vector X containing the *E. coli pUC19* replication origin, the kanamycin resistance gene *nptII*, the yeast  $2\mu$ -replication origin, and a multiple cloning site (MCS). The generated constructs contain the ATFs NLS-JUB1-Gal4AD, NLS-GAL4AD-ANAC102, and NLS-GAL4AD-ATAF1 in combination with two, four, or two copies, respectively, of their cognate binding sites upstream of the minimal *CYC1* promoter. Subsequently, PmeI/AscI-digested plasmid pLOA\_0-1 (28) was used in three separate cloning reactions to assemble: (i) PCR-amplified NLS-JUB1-GAL4AD-2X (primers Plfor/Jubrev, on Entry vector X), and *GDH2*-Ter*GDH2* (primers Gdh2for/Plrev, on yeast BY4741 DNA). The resulting plasmid, in which the expression of *GDH2* is controlled by a JUB1-derived ATF, was called pGNCap01; (ii) PCR-amplified NLS-GAL4AD-ANAC102-4X (primers Plfor/Anacrev, on Entry vector X), and *tHMG1*-Ter*tHMG1* (primers tHMG1for/Plrev, on yeast BY4741 DNA). The resulting plasmid, in which expression of *tHMG1* is controlled by an ANAC102-derived ATF, was called pGNCap02; (iii) PCR-amplified NLS-GAL4AD-ATAF1 (primers Plfor/Atafrev, on Entry vector X), and *ERG20*-Ter*ERG20* (primers ERG20for/Plrev, on yeast BY4741 DNA). The resulting plasmid, in which the expression of *ERG20* is controlled by an ATAF1-derived ATF, was called pGNCap03. Subsequently, *GDH2*-, *tHMG1*-, and *ERG20*-donors (plant-derived ATFs fused to CDSs with left and right HRs to integration sites) were amplified from pGNCap01, pGNCap02, and pGNCap03 vectors using primers Ade2for/Ade2rev, Ura3for/Ura3rev, and His3for/His3rev, respectively. Transformation of the three donors and plasmid pTAJAK105 (29) into strain IMX672.1 (**Table 1**) allowed integration of *GDH2*, *tHMG1*, and *ERG20* into the *ADE2*, *his3D1*, and *ura3-52* loci of the yeast genome, respectively. The newly generated strain was called CaPRedit\_FPP 0.1 (**Table 1**).

### **Construction of reporter plasmids with *ERG9* and *HXT1* promoters**

The yeast *ERG9* and *HXT1* promoters were PCR-amplified from BY4741 genomic DNA (primers ProErg9for/ProErg9rev and ProHxt1for/ProHxt1rev), and inserted into EcoRI/BamHI-digested pGN005B reporter plasmid (16) upstream of the yEGFP coding sequence; the resulting plasmids were called pGN005B-ProERG90 and pGN005B-ProHXT1, respectively. In this way, *ERG9* and *HXT1* promoters drive yEGFP expression (ProERG9-yEGFP and ProHXT1-yEGFP). Reporter plasmids were digested with PmeI and integrated into the *ura3-52* site of yeast strain YPH500, as described (16) to detect their transcriptional activity.

### **Construction of ANAC032-derived synthetic promoters**

To construct synthetic promoters, double-stranded oligonucleotides containing one, two, or four copies of the ANAC032 binding site were inserted into XbaI/SalI-digested reporter plasmid pGN005B (16), upstream of the *CYC1* minimal promoter. Reporter plasmids were linearized with *PmeI* and transformed into yeast strain YPH500 for integration into the genomic *ura3-52* site, as described (16).

### **Construction of plasmids and donor for the downregulation of *ERG9***

To replace the *ERG9* promoter with the synthetic promoter containing the ANAC032 binding site(s), a gRNA targeting a sequence 118-bp upstream of *ERG9* start codon was designed. pCRCT (8) was cut with BsaI and gel-purified to remove 459-bp. The remaining ~10.5-kb plasmid backbone was assembled with annealed single-stranded oligonucleotides (primers gERG9for/gERG9rev), whereby the gRNA sequence was introduced between *Pro*<sub>SNR52</sub> and *Ter*<sub>SUP4</sub>. The resulting plasmid was called pCRCT-ERG9. Moreover, plasmid pGNCap02 was linearized with XbaI. The vector backbone was assembled with annealed single-strand oligonucleotides (primers AnacBSfor/AnacBSrev). Through this, one copy of the ANAC032 binding site was inserted upstream of the *CYC1* minimal promoter. The resulting plasmid was called pGNCap04. Subsequently, *ERG9* donor DNA containing one copy of the ANAC032 binding site upstream of the *CYC1* minimal promoter was amplified from pGNCap04, using primers Erg9for/Erg9rev. To downregulate *ERG9*, *ERG9* donor together with pCRCT-ERG9 is transformed to yeast cells.

### **Construction of plasmid and donors for gene inactivation**

To generate a gRNA for the deletion of *DPP1* CDS, plasmid pCRCT (18) was cut with XhoI and AgeI and gel-purified to remove a 758-bp fragment. The remaining ~10-kb vector backbone was assembled with *Pro*<sub>SNR52</sub> (primers SNR52for/SNR52rev, on pCRCT) and *Ter*<sub>PDC1</sub> (primers PDC1for/PDC1rev, on yeast BY4741 DNA) to generate plasmid pCRCT-DPP1. A gRNA targeting a sequence 395-bp downstream of the *DPP1* start codon, tracrRNA, and *Ter*<sub>SUP4</sub> were introduced between *Pro*<sub>SNR52</sub> and *Ter*<sub>PDC1</sub> by adding them to the primer sequences.

To generate a gRNA for deleting the *LPP1* CDS, pCRCT-DPP1 was cut with SmaI and SalI and gel-purified to remove a 541-bp fragment. The remaining ~10.1-kb vector backbone was assembled with *Ter*<sub>CYC1</sub> (primers CYC1for/CYC1rev, on p426-SNR52p-gRNA.CAN1.Y-SUP4t; Addgene #43803) to generate pCRCT-LPP1. Through this, a gRNA targeting a sequence 432-bp downstream of the *LPP1* start codon and tracrRNA were introduced between *Pro*<sub>SNR52</sub> and *Ter*<sub>SUP4-PDC1</sub>.

To generate gRNA for deleting the *GDH1* CDS, SalI/AscI-digested pCRCT-DPP1 was assembled with double-stranded oligonucleotides (primers GDH1gfor/GDH1grev) introducing

GDH1-gRNA targeting a sequence 664-bp downstream of the start codon. The generated plasmid was called pCRCT-GDH1.

In the next step, pCRCT-DPP1 was cut with NotI and was assembled with PCR-amplified LPP1-gRNA module (primers LPP1for/LPP1rev, on pCRCT-LPP1) introducing LPP1-gRNA expressing module. The generated plasmid called pCRCT-DPP1-LPP1. Then, NotI/PacI digested pCRCT-DPP1-LPP1 was assembled with the PCR-amplified GDH1-gRNA module (primers GDH1for/GDH1rev, on pCRCT-GDH1) introducing the GDH1-gRNA expressing module. The generated plasmid was called pCRCT-GDH1-DPP1-LPP1.

Moreover, three donor DNAs leading to the deletion of *GDH1*, *DPP1*, and *LPP1* were generated by annealing single-strand oligonucleotides (primers gGDH1for/gGDH1rev, gDPP1for/gDPP1rev, and gLPP1for/gLPP1rev, respectively). Each donor contains 50-bp HR to the up- and downstream gene coding sequences allowing deletion of the target gene after CRISPR-mediated DSB. For the deletion of the *GDH1*, *DPP1*, and *LPP1* genes, three donors together with pCRCT-GDH1-DPP1-LPP1 (expressing gRNAs, tracer RNAs, and iCas9 protein) are transformed into yeast cells.

### **Induction experiments, flow cytometry, and data analysis**

To characterize yeast constitutive or ANAC032-derived synthetic promoters, a single colony was grown in 500  $\mu$ l non-inducing SC (synthetic complete) medium containing 2% glucose (w/v) and appropriate selection markers. Cells were grown at 30°C for 36 hours. The main cultures were grown in YPDA medium containing 2% (w/v) glucose (non-inducing medium) or YPDA containing 2% (w/v) galactose and 20 mM IPTG (inducing medium). Cells were grown for 14 - 16 h in a rotary shaker at 30°C and 230 rpm. Samples treated with cycloheximide (500  $\mu$ g/ml) were analysed using a BD FACSCalibur Flow Cytometer (BD Biosciences). The yEGFP fluorescence values were calculated for a minimum of 10,000 cells in each sample as described (16).

### **Strain construction for growth rate measurement by flow cytometry**

To generate constructs suitable for GFP and RFP expression, we inserted a donor DNA containing the *LEU2* gene including its promoter and terminator (primers Leu\_for/Leu\_rev; on pGAD424, TAKARA Bio, GenBank #U07647), the yeast *TEF2* promoter (primers PTEF2\_for/PTEF2\_rev; BY4741 genomic DNA), the AcGFP1 (*Aequorea coerulea* green fluorescent protein, primers GFP\_for/GFP\_rev; plasmid AcGFP1-N1; Addgene #54705) or DsRed (*Discosoma sp.* red fluorescent protein, primers RFP\_for/RFP\_rev; plasmid MSCV-CMV-DsRed-IRES-d24EGFP; Addgene #41944) coding sequences, and the *TEF2* promoter (primers TTEF2\_for/TTEF2\_rev; BY4741 DNA), into the AscI/SbfI digested pLOA\_0\_1 plasmid. The

resulting plasmids, called pNGGreen and pNGRed, were digested with BsaI. Digested pNGGreen was then integrated into the *ura3-52* site of strains IMX672.1, Gen 0.2 and CaPRedit\_FPP 1.0, while digested pNGRed was integrated into the *ura3-52* site of strain IMX672.1. The fluorescent protein reporter strains were called Green\_IMX672.1, Green\_Gen 0.2, Green\_CaPRedit\_FPP 1.0, and Red\_IMX672.1 (WT), respectively (**Table 1**).

A single colony of GFP- and RFP-labeled strains was inoculated in 4 ml YPDA medium containing 2% (w/v) glucose. Cells were grown at 30°C for 36 hours. Growth competitions in single vessels were prepared by mixing GFP- and RFP-labeled cells. We inoculated co-cultures at a ratio of 9 : 1 (optimized : WT) at OD<sub>600</sub> of ~0.1 in 4 ml YPDA medium. Cells were grown for 48 h in a rotary shaker at 30°C and 230 rpm. At time points 6, 24, and 48 h, samples were removed for OD<sub>600</sub> measurement and analysis by flow cytometry to determine the ratio of GFP-positive to RFP-positive cells. At each time point, cells were diluted 100-fold into fresh liquid medium for growth until the next time point. We counted RFP- and GFP-labeled cells.

### **Integration of foreign $\beta$ -carotene pathway genes into the yeast genome**

The module containing *Mcr1*-, *Mcr1YB*-, and *BTS1* CDSs of the  $\beta$ -carotene pathway downstream of three plant-derived ATFs, namely NLS-GAL4AD-GRF7/4X, NLS-ATAF1-GAL4AD/2X, and NLS-GAL4AD-RAV1/4X, respectively (see **Supplementary Methods**) were integrated into the *X-3*, *XI-2*, or *XII-5* loci, respectively, of yeast strains with a CEN.PK background (IMX672.1, Gen 0.2 and CaPRedit\_FPP 1.0). To simultaneously integrate three genes required for  $\beta$ -carotene production, strains were co-transformed with either 1  $\mu$ g of triple gRNA plasmid pCOM04 (see **Supplementary Methods**) plus 1  $\mu$ g of each donor fragment: *Mcr1* (primers DES1X3\_for/DES1X3\_rev, on Destination vector I-Mcr1), *BTS1* (primers ACCEPTAXII5\_for/ACCEPTA-XII5\_rev on Acceptor vector A-BTS1), *Mcr1YB* (primers ACCEPB-XI3for/ACCEPTB-XI3rev on Acceptor vector B-Mcr1YB), plus 1  $\mu$ g of triple gRNA plasmid pCOM004. Thereby, *Mcr1*, *BTS1*, and *Mcr1YB* were integrated into the *X-3*, *XI-2*, or *XII-5* locus, respectively. Cells were plated on SC-Leu/-Ura/-His/-Lys medium to screen for gRNA plasmid which harbours the *LYS2* selection marker and  $\beta$ -carotene producing colonies with integrated selection markers. When colonies appeared, the transformation plates were replicated on non-selective induction plates (YPDA, 2% (w/v) galactose, 20  $\mu$ M IPTG).

### **High-performance liquid chromatography (HPLC)**

Yeast colonies were inoculated into 4 ml non-inducing medium (SC with appropriate selection marker) and grown for 18 - 24 h at 30°C and 230 rpm in a rotary shaker. Subsequently, the pre-cultures were used to inoculate main cultures (50 ml) in inducing medium (SC with the

appropriate selection marker and 20 mM IPTG, 2% (w/v) galactose). All shake-flask cultures were inoculated from pre-cultures to an initial OD<sub>600</sub> of 0.1. Cells were then grown in 500 ml shake flask for 3 days at 30°C and 230 rpm to saturation. Next, cells were harvested for carotenoid analysis. Carotenoid extraction was carried out according to the acetone extraction method with some modifications (31). The cell pellet was washed once with deionized water. Glass beads (400 - 600 µm diameter; Sigma Chemicals) and 500 µl of acetone were added to the cell pellet in a 2.0-ml microcentrifuge tube (Eppendorf) and vortexed to pulverize the cells, followed by incubation for 20 min at 30°C. After breakage, the bead-cell mixture was centrifuged at 13,000 rpm in a microcentrifuge for 5 min at 4°C, and the acetone supernatant was collected in 2-ml microcentrifuge tubes. This extraction procedure was repeated until the cell pellet was white. The combined acetone extracts were transferred to a glass vial and dried using a speed-vac. To avoid degradation of carotenoids by light, the glass tubes were kept inside a black colored 1.5-ml microcentrifuge tubes (Eppendorf). Moreover, all procedures were performed under green save light. The dry samples were stored at -80°C and solved in acetone before HPLC measurement. HPLC analyses were carried out in triplicate by AppliChrom (Oranienburg, Germany). Carotenoids were separated by HPLC using a RP-HPLC phase AppliChrom OTU DiViDo (250 x 4.6 mm) column with porous 5-µm particles (85/10/5, acetone/methanol/isopropanol, v/v/v) as the mobile phase, with a 1.0-ml/min flux. The elution profiles were recorded using Shimadzu 450 nm (D2) and Kontron 300 nm (D2) detectors.

## RESULTS

### Rationale and design of the CaPRedit system for enhancing FPP production

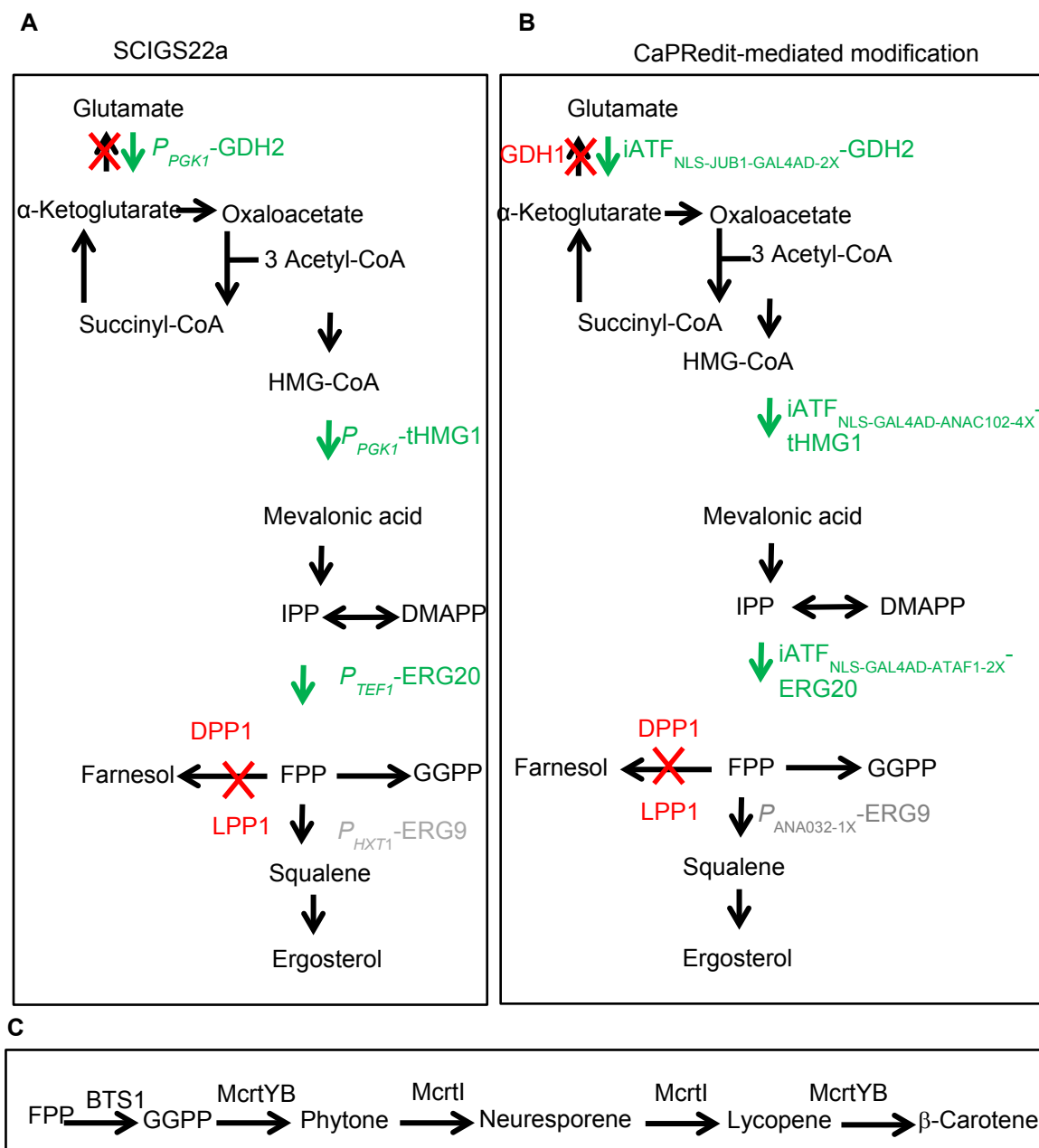
In this study, we applied CRISPR/Cas9-facilitated multi-locus integration of plant-derived regulators to inducibly overexpress desired genes in the *S. cerevisiae* genome by combining three recently reported tools: (i) *in vivo* DNA assembly, (32) (ii) homology-directed repair of DSBs using CRISPR/Cas9 (33) for marker-free, multi-locus genome engineering, and (iii) our recently reported new class of inducible ATFs to highly express core metabolic pathway genes. (16) We named our method CaPRedit for 'CRISPR/Cas9 and Plant-derived Regulator-mediated genome **editing**'.

As an application of CaPRedit, we planned to construct a *S. cerevisiae* cell factory for isoprenoid production by redirecting carbon towards the production of FPP and minimizing FPP consumption for the metabolites that compete with β-carotene synthesis from FPP through three types of modifications (**Figure 1**): (i) overexpression of *GDH2*, *tHMG1*, and *ERG20* under the control of an inducible plant-derived ATF. Thereby, an extra copy of each gene is introduced into

the yeast genome adding to the native, endogenous genes; (ii) downregulation of *ERG9* expression by replacing the promoter of the endogenous *ERG9* gene (*Pro<sub>ERG9</sub>*) with a weaker, heterologous promoter; and (iii) partial deletions, leading to frame shift mutations, of the *DPP1*, *LPP1* and *GDH1* genes. To alter expression *GDH2*, *tHMG1*, and *ERG20*, constitutive promoters were previously implemented to generate strain SCIGS22a (**Figure 1A**), while CaPRedit uses plant-derived regulators (**Figure 1B**). We named the new overexpression strain CaPReditFPP 1.0. Subsequently, accumulation of  $\beta$ -carotene was compared in CaPRedit\_FPP 1.0, IMX672.1 and Gen 0.2 (**Table 1**). Later, after introducing all modifications including the downregulation of *ERG9* expression and the inactivation of the *DPP1*, *LPP1* and *GDH1* genes into strain CaPRedit\_FPP 1.0,  $\beta$ -carotene production will be measured in the CaPRedit-derived strain.

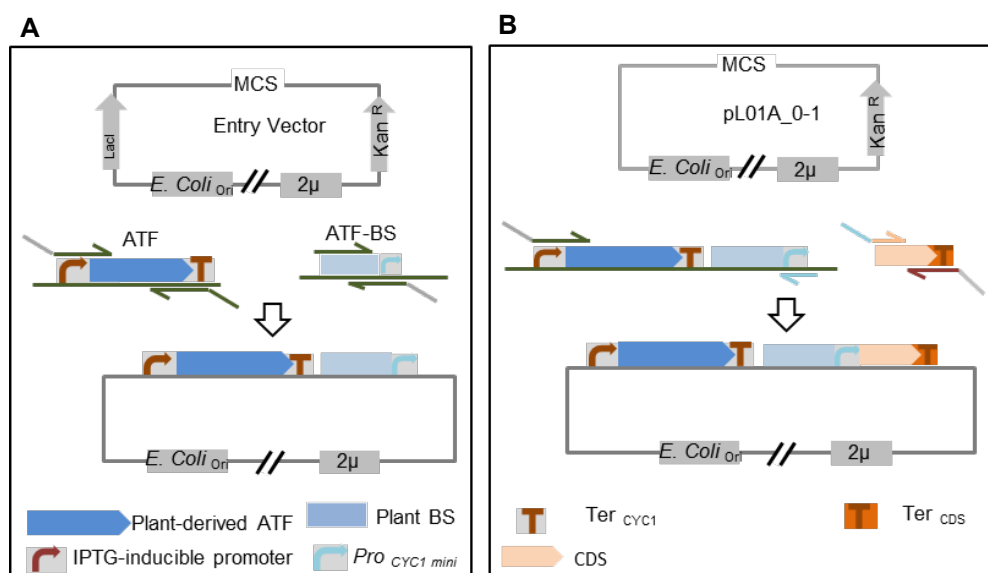
### **General strategy for designing units of plant-derived ATFs and pathway genes**

For each expression donor of the CaPRedit system, two modules are needed: an ATF and a CDS fused to its native terminator. To generate ATF module, the inducible plant-derived ATF amplified (*Pro<sub>mGAL1-IPTG</sub>-plant-derived ATF-Ter<sub>CYC1</sub>*) from expression plasmid (16) and the synthetic promoter containing corresponding plant TF binding site (*Pro<sub>miniCYC1-BS</sub>*) from the reporter plasmid (16) were assembled in the Entry vector X (**Figure 2A**, see **MATERIALS AND METHODS**) and it then used to PCR-amplify the *Pro<sub>mGAL1-IPTG</sub>-plant-derivedATF-Ter<sub>CYC1</sub>-Pro<sub>miniCYC1-BS</sub>*. In the next step, the CDS of the gene of interest and its downstream terminator were PCR-amplified from the yeast genome and together with the ATF module assembled in plasmid pLOA\_0\_1 (28), as shown in **Figure 2B**.



**Figure 1. Engineering platform for FPP precursor production in yeast.** To redirect yeast metabolisms toward the FPP production, GDH2, ERG20, and tHMG1 overexpression (green), ERG9 downregulation (grey), inactive or deleted DPP1, LPP1 and GDH1 (red) are needed. (A) Schematic illustration of SCIGS22a. The yeast constitutive promoters are used to change the expression level of enzymes. (B) Schematic illustration of strain optimized by CaPRedit approach. The inducible plant-derived ATFPs are used to change the expression level of enzymes. (C)  $\beta$ -Carotene production from FPP precursor. The green, red, and grey arrows/genes indicate the expressed, deleted, and downregulated genes respectively. NLS: nuclear localization signal from the SV40 large T antigen; JUB1: plant JUNGBRUNNEN1 transcription factor; GAL4AD: yeast GAL4 activation domain; ANAC102: NAC domain

containing protein 102; ATAF1: *Arabidopsis thaliana* Activating Factor 1; ANAC032: NAC domain containing protein 32; '1X', '2X', and '4X' indicate one, two, or four copies of the TF binding site, respectively; ERG20: FPP synthase; DPP1 and LPP1: lipid phosphate phosphatases; ERG9: squalene synthase; tHMG1: truncated HMG-CoA reductase, GDH1: NADP<sup>+</sup>-glutamate dehydrogenase; GDH2: NAD<sup>+</sup>-dependent glutamate dehydrogenase; BTS1: geranylgeranyl diphosphate synthase; MctYB: optimized phytoene synthase/lycopene cyclase; MctI: phytoene desaturase; HMG-CoA: 3-hydroxy-3-methylglutaryl-CoA; IPP: isopentenyl diphosphate; DMAPP: dimethylallyl diphosphate; FPP: farnesyl diphosphate; GGPP: geranylgeranyl diphosphate.



**Figure 2. Construction of overexpression donors.** (A) Construction of plant-derived ATFP units. Coding sequence of plant-derived ATFs under the control of an IPTG inducible promoter and their cognate binding sites located upstream of the yeast *CYC1* minimal promoter were obtained by PCR using appropriate expression and reporter plasmids respectively.(16) Subsequently, Entry vector X (see **Supplementary Methods**) digested at *MCS*, and were then reassembled with two fragments through homologous-based cloning method. (B) Construction of plasmid delivering overexpression donors. The CDS of yeast endogenous gene fused to its downstream terminator and plant-derived ATFP from Entry vector X were PCR-amplified and subsequently were assembled in *PmeI/AscI*-digested pL0A\_0\_1. *MCS*: multiple cloning site; *E. coli Ori*: *pUC19* replication origin of *E. coli*; *KanR*: kanamycin resistance gene *nptII*; *LacI*: lactose inhibitor; *2μ*: yeast two-micron replication origin.

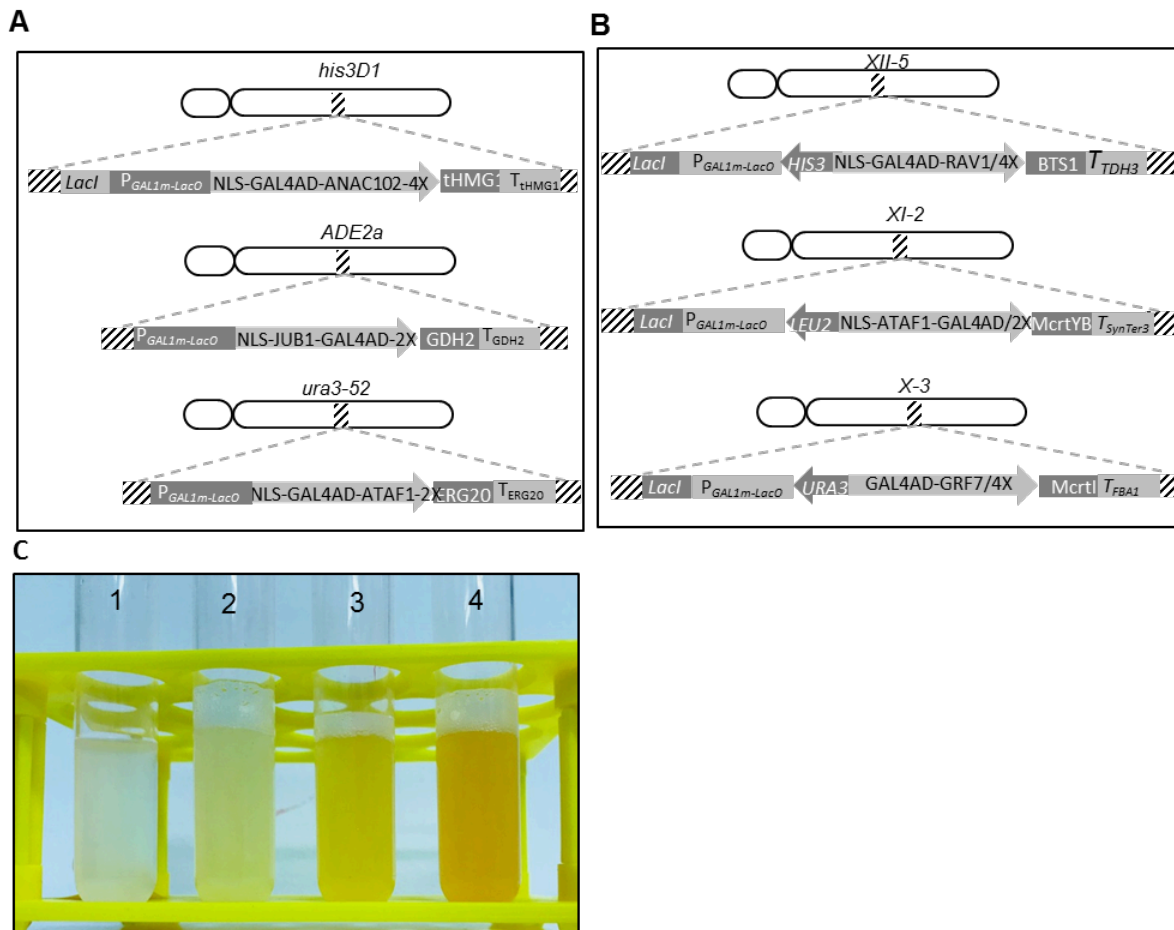
### CaPRedit-mediated one-step genome integration of multiple ATF units and genes

We used one-step multiple CRISPR/Cas9-mediated genome editing, implemented in CaPRedit, to overexpress the *tHMG1*, *ERG20*, and *GDH2* genes. To this end, units of ATFs and pathway genes are used as donor fragments for integration into the respective loci. Moreover, each donor



also contains 50-bp long overhang sequences at their 5'- and 3' ends to facilitate integration into the pre-designed genomic loci. To trigger  $\beta$ -carotene production, the three pathway genes *GDH2*, *tHMG1*, and *ERG20* were overexpressed in yeast. To this end, the NLS-JUB1-GAL4AD-2X-GDH2-Ter<sub>GDH2</sub>, NLS-GAL4AD-ANAC102-4X-tHMG1-Ter<sub>tHMG1</sub>, and NLS-GAL4AD-ATAF1-ERG20-Ter<sub>ERG20</sub> donors, amplified from plasmids pGNCap01, pGNCap02, and pGNCap03 (see **MATERIALS AND METHODS**), were integrated into the *ADE2.a*, *his3D1*, or *ura3-52* locus, respectively, which were previously characterized as high-efficiency integration sites (8). To this end, plasmid pTAJAK105 (8) expressing gRNAs targeting the *ADE2.a*, *his3D1*, and *ura3-52* loci, the Cas9-expressing plasmid pCRCT (9), and DNA donor fragments encoding ATFs and pathway enzymes were transformed into IMX672.1 (**Figure 3A**, **Supplementary Figure S2**). The successfully constructed strain was named CaPRedit 1.0. Cells were plated on media that select for the presence of the gRNA and Cas9 plasmids, which harbor the *LEU* and *URA* selection markers, respectively (SC-Leu-Ura). When colonies appeared, successfully mutated colonies were confirmed by sequencing PCR-amplified fragments containing the ATF and the pathway genes (NLS-JUB1-GAL4AD-2X-GDH2-Ter<sub>GDH2</sub>, primers ATFfor/GDH2rev; NLS-GAL4AD-ANAC102-4X-tHMG1-Ter<sub>tHMG1</sub>, primers ATFfor/tHMG1rev; NLS-GAL4AD-ATAF1-ERG20-Ter<sub>ERG20</sub>, primers ATFfor/ERG20rev). In addition, colonies were replicated on non-selective media plates (YPDA medium) to eliminate the gRNA- and Cas9-expressing plasmids. Here, strain CaPRedit\_FPP1.0 was generated which overexpresses *GDH2*, *tHMG1*, and *ERG20* to serve as a parental strain for the generation of strains with downregulated *ERG9* expression and inactivated *GDH1*, *DPP1*, and *LPP1* genes. We indirectly measured the intracellular levels of FPP by assaying  $\beta$ -carotene accumulation after the co-expression of genes encoding McrtI, McrtYB, and BTS1 leading to the conversion of FPP to  $\beta$ -carotene. To this end, the donor cassettes containing these  $\beta$ -carotene pathway genes and their upstream ATFs (NLS-GAL4AD-GRF7/4X\_McrtI, NLS-ATAF1-GAL4AD/2X\_McrtYB, and NLS-GAL4AD-RAV1/4X\_BTS1) (**Supplementary Figure S2**) were integrated into the *X-3*, *XI-2*, and *XII-5* loci, representing previously characterized high-efficiency integration sites (30), of strains Gen 0.2, IMX672.1, and CaPRedit\_FPP 1.0 (**Figure 3B**). As the *URA3*, *LEU2*, and *HIS3* genes are encoded on the assembly modules, the selection for successful assembly of the parts was performed by plating the yeast cells on medium containing auxotrophic selection markers (SC-Ura/-Leu/-His). Orange yeast colonies were observed on the selection plates, and when single orange-colored colonies were re-streaked on YPDA medium, all resulting colonies were of uniform color (**Supplementary Figure S3**), indicating robust gene expression from the assembled pathways. The strain containing the ATF-CDSs of the  $\beta$ -carotene pathway in the CaPRedit\_FPP 1.0

background showed a qualitatively stronger  $\beta$ -carotene accumulation than strains established in the Gen 0.2 and IMX672.1 backgrounds (**Figure 3C**). We selected three colonies from each strain for quantitative determination of  $\beta$ -carotene content by HPLC. The results shown in **Table 2** demonstrate that overexpression of the three genes *Mcrt1*, *McrtYB*, and *BTS1* in CaPRedit\_FPP 1.0 leads to 4.3- and 1.3-fold higher accumulation of  $\beta$ -carotene ( $0.61 \pm 0.04$  mg  $\beta$ -carotene/g (dw)) than in strains IMX672.1 ( $0.14 \pm 0.047$  mg  $\beta$ -carotene/g (dw)) and Gen 0.2 ( $0.45 \pm 0.05$   $\beta$ -carotene/g (dw)).



**Figure 3.  $\beta$ -Carotene production in strain CapRedit\_FPP 1.0.** (A) Scheme showing donors derives CaPRedit\_FPP 1.0 strain. ATFPs NLS-GAL4AD-ANAC102-4X, NLS-JUB1-GAL4AD-2X, and NLS-GAL4AD-ATAF1-2X were fused to tHMG1-TerHMG1, GDH2-TerGDH2, and ERG20-TerERG20. The tHMG1, GDH2, and ERG20 donors are flanked by 50-bp homology arms to integrate into the *his3D1*, *ADE2.a*, and *ura3-52* loci. In each donor, modified GAL1 promoter is located upstream of plant-derived ATF which contains *lacO* site. Additionally, the tHMG1 donor expresses *LacI*. The corresponding binding site of plant-derived ATFs is placed upstream of *CYC1* minimal promoter to drive its downstream gene production. (B) Scheme showing donors derive  $\beta$ -carotene production. *Mcrt1*-, *BTS1*- and *McrtYB*-CDS donors are flanked by homology arms to integrate into the *X-3*, *XI-2*, and *XII-5* loci respectively. Each

donor contains LacI repressor, the IPTG inducible GAL1 modified promoter with *lacO* site upstream of plant-derived ATFPs (see **Supplementary Figure S2**). Selection on SC-Ura/-Leu/-His media allows screening for successfully integrated donors. (C). Representative liquid culture of the constructed strains after plant-derived ATF induction (IPTG and galactose). Colors of different carotenoid-producing *S. cerevisiae* strains in YPDA media for wild type (1), IMX672.1 (2), Gen 0.2 (3), and CaPRedit\_FPP 1.0 (4) strains. To simplify the figure, the *CYC1* terminator located downstream of plant-derived ATFs is not.

**Table 2.** HPLC analysis of carotenoid content in engineered yeast strains.

| Strain                               | $\beta$ -Carotene (m $\mu$ g $^{-1}$ cdw) |
|--------------------------------------|---|
| IMX672.1                             | 0.007 $\pm$ 0.001                         |
| Gen 0.2                              | 0.015 $\pm$ 0.003                         |
| IMX672.1 + $\beta$ -carotene         | 0.14 $\pm$ 0.047                          |
| Gen 0.2 + $\beta$ -carotene          | 0.46 $\pm$ 0.05                           |
| CaPRedit_FPP 1.0 + $\beta$ -carotene | 0.61 $\pm$ 0.044                          |

Values represent the mean of three independent colonies after three days of cultivation. Cdw, cell dry weight.

In the present study, we tested the effect of overexpressing *GDH2*, *tHMG1*, and *ERG20* on FPP production; overaccumulation of FPP in strain CaPRedit\_FPP 1.0 was attributed to an improved production of  $\beta$ -carotene. We plan to quantify the expression of *GDH2*, *tHMG1*, and *ERG20* (encoding enzymes that enhance metabolite flux towards FPP), *BTS1*, *McrYB*, and *McrI* (encoding enzymes that convert FPP to  $\beta$ -carotene), and *ERG9* (encoding the enzyme which converts FPP to squalene) in CaPRedit 1.0, Gen 0.2 and IMX672.1 by carrying out quantitative real-time reverse-transcription PCR (qRT-PCR).

### Growth rate measurements

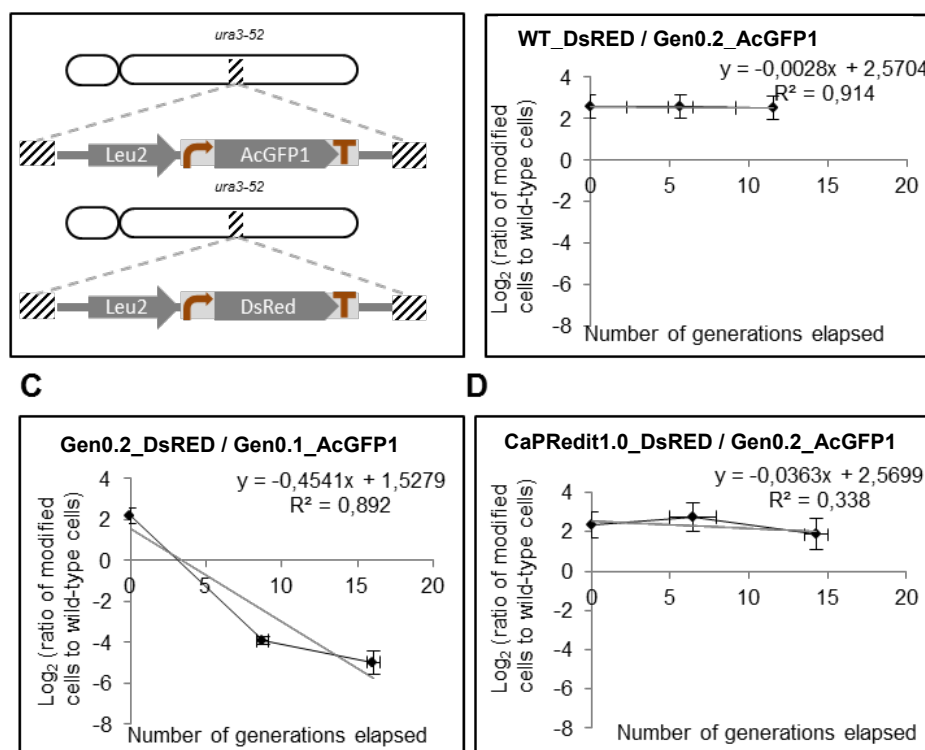
In order to assess whether a growth penalty is associated with *GDH2*, *tHMG1*, and *ERG20* overexpression in the engineered strains, we implemented a previously developed high-throughput fitness assay in which flow cytometry is used to monitor growth competition of fluorescently labeled strains (34). In this assay, AcGFP1-expressing wild-type and RFP-expressing modified strains were co-cultured. The co-cultures maintained were carried by serial dilution. At each time point, samples were removed for analysis by flow cytometry to determine the ratio of AcGFP1-positive to DsRED-positive cells. Wild-type and modified cell populations are resolved by flow cytometry. Rare events that appear to be both DsRED-positive and AcGFP1-positive represent instances in which a modified cell and a wild-type cell are

misidentified by the cytometer as a single cell, and we took this into account during analysis. (**Figure 4A**) Because a large number of individual cells (20,000) can be measured by flow cytometry, this assay makes it possible to calculate relative growth rates by determining the rate of change of the ratio of modified strain cells to WT cells over the course of the competition (34).

While growth of IMX672.1 is not significantly affected by the expression of AcGFP1 or DsRED fluorescence proteins (growth rate of IMX672.1 + DsRED relative to IMX672.1 + AcGFP1 is  $\sim -0.003$ , **Figure 4B**), we found that the Gen 0.2 strain (SCGSSa background) exhibited a major growth defect (growth rate of Gen 0.2 + DsRED relative to IMX672.1 + AcGFP1 is  $\sim -0.45$ , **Figure 4C**). In contrast, we detected only a very small growth defect in the CaPRedit\_FPP 1.0 strain (growth rate of CaPRedit\_FPP 1.0 + DsRED relative to IMX672.1 + AcGFP1 is  $\sim -0.03$ , **Figure 4D**). Our results demonstrated that reintroducing three yeast native enzymes under the control of strong, inducible ATFPs has only a slight negative effect on yeast growth (in the absence of the inducer, *i.e.* ATFs and enzymes were not expressed, except leaky expression). In contrast, constitutive expression of the same enzymes using a relatively weak, constitutive yeast promoter in combination with one downregulated enzyme and the deletion of three more enzymes has a strong growth defect. Although CaPRedit\_FPP 1.0 strain does not contain the downregulation and deletion modifications, and thereby we have not studied their effect on yeast growth here, one strong possibility is that the lack of a pronounced growth defect in the CaPRedit\_FPP 1.0 strain is due to the inducible rather than constitutive overexpression of pathway enzymes. In order to test this, we plan to introduce the same Gen 0.2 deletions and downregulation into CaPRedit\_FPP 1.0.

We assessed the accuracy of the flow cytometry–based technique by measuring the growth of strains in triplicate biological replicate measurements were in better agreement for Gen 1.0 strain (correlation coefficient squared between logarithm base 2 of number of modified to wild type cells and number of elapsed generation variables ( $R^2$ ):  $\sim 0.89$ ; **Figure 4C**) than for CaPRedit\_FPP 1.0 strain ( $R^2$ :  $\sim 0.34$ ; **Figure 4D**).

**A****B**



**Figure 4. Growth rate measurements.** (A) Schematic of the fluorescently-labelled strains. Fluorescence expressing plasmids harboring the *LEU2* gene as selection marker, either AcGFP1 or DsRed under the control of the *TEF2* promoter are integrated into the *ura3-52* site of yeast genome. Relative growth rates of yeast wild type (B), Gen 0.2 (C), and CaPRedit\_FPP 1.0 (D). AcGFP1-expressing wild-type strain is co-cultured with a RFP-expressing strains in ratio 1:9, and the relative abundance of each strain is monitored over time by flow cytometry. Gray line indicates the rate of RFP-labelled strain depletion over time. Full data of **Figure 4B**, **4C**, and **4D** are shown in **Supplementary Data S1A**, **S1B**, and **S1C**.

### Selection of synthetic promoters for downregulation of *ERG9* expression

CaPRedit employs constitutively expressed plant-derived transcriptional regulators with lower transcriptional output than the native promoters of genes that are object of downregulation. To minimize overflow to sterols that use FPP as a precursor, and thereby to maximize redirection of FPP toward GGPP production (**Figure 1**), the native *ERG9* promoter ( $Pro_{ERG9}$ ) needs to be replaced with a weaker, constitutively expressed promoter. Previously, a glucose-sensing promoter ( $Pro_{HXT1}$ ) was used for this purpose (5,36). Here we plan to implement a plant-derived synthetic promoter with a transcriptional output similar to that of  $Pro_{HXT1}$ . Implementing plant-derived promoters will provide metabolic engineering projects with more choices, when a large collection of weak promoters is needed.

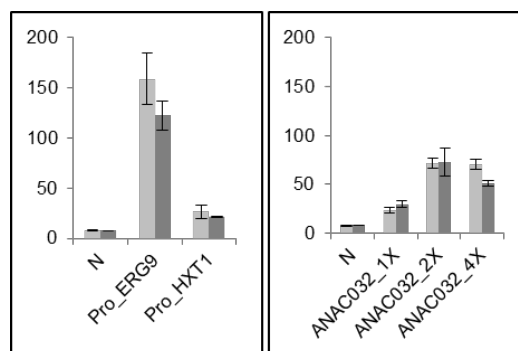
To choose the plant-derived promoter with transcriptional output similar to that of  $Pro_{HXT1}$ , the activities of the  $Pro_{ERG9}$  and  $Pro_{HXT1}$  promoters were tested using our previously developed

reporter system (16). The system relies on yeast enhanced green fluorescent protein (yEGFP) expression controlled by an upstream promoter. The reporter gene is integrated into the *ura3-52* locus of the genome of yeast YPH500 strain and fluorescence output is measured. More specifically, we inserted the *Pro<sub>ERG9</sub>* and *Pro<sub>HXT1</sub>* sequences upstream of the *yEGFP* coding sequence. Our data demonstrated that *Pro<sub>HXT1</sub>* is 6-fold weaker than *Pro<sub>ERG9</sub>* (**Figure 5A**).

Previously, we studied the orthogonality of plant TF binding sites using the reporter system (16). We tested to what extent endogenous yeast TFs can activate expression of the yEGFP reporter through interaction with two copies of *cis*-regulatory elements of different plant TFs, in the absence of the plant TF (16). We expected to observe zero to very low expression for binding site of plant TFs (16). Surprisingly, some degree of basal expression was observed for the binding motifs of some plant TFs, indicating the endogenous yeast transcriptional machinery can interact with certain plant *cis*-regulatory elements. Although such plant *cis*-regulatory elements are not appropriate candidates for orthogonal regulation of gene expression in the yeast *S. cerevisiae*, they can be a suitable alternative to yeast native constitutive promoters that are commonly used in synthetic biology projects. The test result showed that two copies of the ANAC102 binding site that resulted in higher basal expression than other tested binding sites is 3.5-fold weaker than *Pro<sub>ERG9</sub>* as determined by the reporter system (16).

To achieve a synthetic promoter containing ANAC032 binding site(s) with a transcriptional output similar to that of *Pro<sub>HXT1</sub>*, we tested the transcriptional output of a synthetic promoter containing ANAC032 binding site in one, two, and four copies. As shown in **Figure 5B**, *Pro<sub>CYC1min\_ANAC032-1X</sub>* and *Pro<sub>HXT1</sub>* resulted in a similar fluorescence output. Therefore, *Pro<sub>CYC1min\_ANAC032-1X</sub>* can be used to downregulate *ERG9* gene expression.

**A****B**



**Figure 5. Characterization of promoter for downregulation.** (A) Characterization of yeast constitutive promoters. *ERG9* and *HXT1* promoter sequences cloned from the yeast genome give consistent expression of yEGFP fluorescent reporters. (B) Binding site specificity of ANAC032 in yeast. One, two, and four copies ('1X', '2X', '4X') of binding sites of ANAC032 were assembled into plasmid pGN005B (16) upstream of the *CYC1* minimal promoter to control yEGFP expression. Next, the resulting plasmids were integrated into the *ura3-52* locus of the genome of yeast YPH500 strain and fluorescence output was measured. N, empty pGN005B (*CYC1* minimal promoter-derived yEGFP expression). Data are geometric means  $\pm$  SD of the fluorescence intensity obtained from three cultures, each derived from an independent yeast colony and determined in three technical replicates. AU, arbitrary units. Full data are shown in **Supplementary Data S2**.

## DISCUSSION

To increase the product titers for commercialization, the genome of the host must be engineered toward enhancing production of the precursors, and the expression heterologous engineered pathway need to be balanced (37). CaPRedit present a promising tool to shorten the gap between genome engineering pathway engineering, as it was developed with two goals in mind: to speed up strain modification and to facilitate very low to high expression of key enzymes. The system presented here relies upon, CRISPR/Cas9-mediated one-step multigene modification system (8) and inducible plant-derived regulators with wide range of expression (weak to super-strong) (19).

A wider range of transcriptional activation capacities are required for metabolic engineering projects (2). Moreover, individual controlling module is required for expression of each gene in eukaryotic organisms (*e.g.*, yeast), because eukaryotic transcription does not employ polycistronic messenger RNAs (38). A small set of well-characterized of yeast inducible and constitutive promoters have been extensively characterized and are commonly used for synthetic biology projects. Nevertheless, recognition of inadequacies in these often-used promoters has led to the development of libraries of ATFs, such as TAL-, CRISPR-based regulators (17), and plant-derived ATFs (16). In the current study, we achieved one-step

genomic integration of synthetic modules containing IPTG-inducible ATFs via CRISPR/Cas9 double-strand breaks.

To engineer the complex metabolic pathways, production of several targets often need to change in different ways through increasing the expression of key enzymes, decreasing the expression of necessary genes and removing the expression of genes involved in competing pathways (37). CaPRedit can be implemented to (i) introduce the heterologous enzymes involved in a biosynthetic pathway engineered in yeast and (ii) to redirect yeast *S. cerevisiae* endogenous metabolic flux toward a key precursor of a desired heterologous product via down- and overexpression of genes. In general, to do the overexpression group of modifications, either the units of inducible plant-derived ATFs together with pathway genes (to introduce another copy of gene) or units of plant-derived ATFs (to replace the original promoter of gene) are used as donor fragments to integrate into the wanted loci, whereas to do downregulation group of modifications, the native promoter of genes are replaced with weaker constitutively overexpressed plant-derived regulators.

As a test case for the CaPRedit system, we engineered yeast to produce significantly more FPP. FPP is a central feed-back regulator in the MVA pathway (5) (**Figure 1**) and is a precursor of several different isoprenoid products and squalene (39). Production of non-native isoprenoid carotenoids is also derived from the same FPP. Numerous studies have reported increased production of carotenoids in yeast through engineering the heterologous pathway consuming the yeast endogenous FPP precursor (18,21). In contrast, there are few reports focusing on redirecting the yeast endogenous metabolic flux toward the production of FPP through overexpression of enzymes involved in FPP production and then redirecting FPP toward  $\beta$ -carotene production through deletion, and/or downregulation of enzymes involved in pathways that compete with  $\beta$ -carotene synthesis for FPP (5).

We reported upon only inducibly overexpressed *GDH2*, *tHMG1*, and *ERG20* using three strong plant-derived ATFPs (2.5- to 5.5-fold stronger than yeast *TDH3* promoter), we constructed *S. cerevisiae* strain producing  $0.61 \pm 0.044$  mg  $\beta$ -carotene/g (dw), which is 1.3-fold more than the previously optimized strain (that harbour modifications lead to reduction in FPP consumption by competing pathways, in addition to constitutively overexpressed *GDH2*, *tHMG1*, and *ERG20* genes (5,18). To produce  $\beta$ -carotene, a combination of weak, medium, and strong (IPTG) inducible plant-derived ATFPs (based on our personal observation, unpublished data) was used to express the *BTS1*, *McrtI*, and *McrtYB* genes, converting FPP to  $\beta$ -carotene.



To succeed in generating yeast cell factories, metabolic engineering projects require thoughtful design. The most efficient designs typically incorporate a biomass production phase followed by a target production phase. By utilizing an inducible promoter to control expression of strong plant-derived ATFPs in CaPRedit\_FPP 1.0, expression of ATFs, and any potential interactions with the host genome and proteins that impose fitness costs can be delayed until target production is desired. We measured differential growth of IMX672.1 (wide-type cells), Gen 0.2 (modified cells with deleted, constitutively overexpressed and downregulated genes leading to enhanced production of FPP), and CaPRedit\_FPP 1.0 strain (modified cells with overexpressed genes under the control of IPTG-inducible ATFPs leading to enhanced production of FPP) while ATF expression is prevented in yeast log and lag phase of growth (CaPRedit\_FPP 1.0 strain). Approaches to quantitate growth that are based on colony size or optical density measurements are affected by the growth microenvironment, and thereby external sources of variation affect true biological differences. Methods relying on competitive growth can overcome these issues by measuring relative growth differences among strains competing in a homogeneous environment (38). Here, we used a previously reported flow cytometry–based high-throughput technique (34) that relies on fluorescent protein expression to monitor cell number and thereby growth rate. We observed that the use of constitutive promoters has a major negative effect on yeast growth. We observed little growth defect for CaPRedit\_FPP 1.0 strain in conditions where the plant-derived ATFPs were not induced and therefore *GDH2*, *tHMG1*, and *ERG20* genes were not overexpressed. This emphasizes that utilizing inducible promoters for expression of ATFPs may minimize consumption of cell energy and nutrient resources and maximize biomass production capacity prior to the production phase. As a likely consequence of the separation of these two phases, 1.3-fold more  $\beta$ -carotene was produced in CaPRedit\_FPP 1.0 strain than in Gen 0.2 strain.

To construct an even more efficient *S. cerevisiae* cell factory for isoprenoid production from FPP, we plan to (i) partially delete *DPP1*, *LPP1*, and *GDH1* genes, and (ii) integrate the binding site of plant TF of upstream of *ERG9* to lower its expression level. Furthermore, we plan to investigate the expression level of the FPP gene in CaPRedit\_FPP, Gen 0.2, and IMX672.1 strains using qRT-PCR.

Since IPTG is expensive, and induction is irreversible, it is not an ideal choice for industrial scale production. However, the CaPRedit system is compatible with other chemically inducible promoters and light-controlled molecular switches may be the most promising alternative for industrial applications (40).

The CaPRedit strategy presented here could be extended to construct other pathways in *S. cerevisiae* to increase production of endogenous precursor(s) and/or modify heterologous

enzyme expression in a metabolic pathway. We have demonstrated our CaPRedit approach to (i) be a powerful tool to utilize synchronized, inducible, strong plant-derived ATFPs that are 8- to 10-fold stronger than the yeast *TDH3* promoter and to (ii) be a fast genome editing system due to the highly efficient CRISPR/Cas9 system that allows one-step, multiloci targeted genomic integration.

## ACKNOWLEDGEMENTS

We thank Yansheng Zhang (Chinese Academy of Sciences, Beijing, China) for McrtI and McrtYB; Jules Beekwilder (Plant Research International, Wageningen, Netherlands) for yeast IMC167 strain; Michael K. Jensen (Technical University of Denmark, Kongens Lyngby, Denmark) for pTAJAK-105; Alex T. Nielsen (Technical University of Denmark, Lyngby, Denmark) for pTAJAK-92; and Verena Siewers (Chalmers University of Technology, Göteborg, Sweden) for yeast strain SCIGS22a. I am also thankful to Karina Schulz for helping me to construct the plasmids.

## REFERENCES

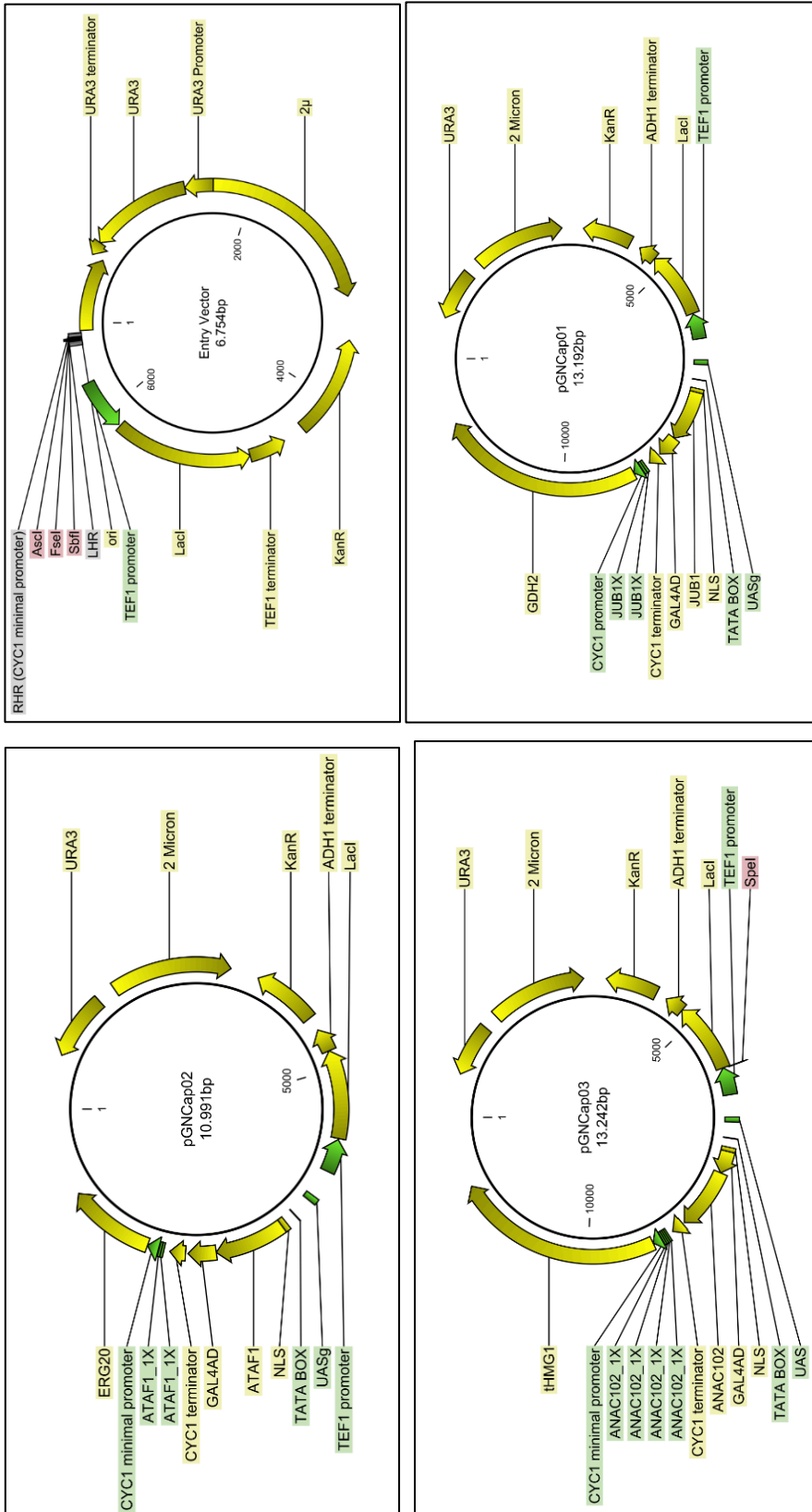
1. Hong, K.K. and Nielsen, J. (2012) Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell Mol Life Sci*, 69, 2671-2690.
2. Generoso, W.C., Schadeweg, V., Oreb, M. and Boles, E. (2015) Metabolic engineering of *Saccharomyces cerevisiae* for production of butanol isomers. *Curr Opin Biotechnol*, 33, 1-7.
3. Saha, N., Samanta, A.K., Chaudhuri, S. and Dutta, D. (2015) Characterization and antioxidant potential of a carotenoid from a newly isolated yeast. *Food Science and Biotechnology*, 24, 117-124.
4. Jarboe, L.R., Zhang, X., Wang, X., Moore, J.C., Shanmugam, K.T. and Ingram, L.O. (2010) Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. *J Biomed Biotechnol*, 2010, 761042.
5. Scalcinati, G., Partow, S., Siewers, V., Schalk, M., Daviet, L. and Nielsen, J. (2012) Combined metabolic engineering of precursor and co-factor supply to increase alpha-santalene production by *Saccharomyces cerevisiae*. *Microb Cell Fact*, 11, 117.
6. Futcher, B. and Carbon, J. (1986) Toxic Effects of Excess Cloned Centromeres. *Molecular and cellular biology* 6, 2213-2222.
7. Tyo, K.E., Ajikumar, P.K. and Stephanopoulos, G. (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat Biotechnol*, 27, 760-765.
8. 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 Synth Biol*, 4, 585-594.

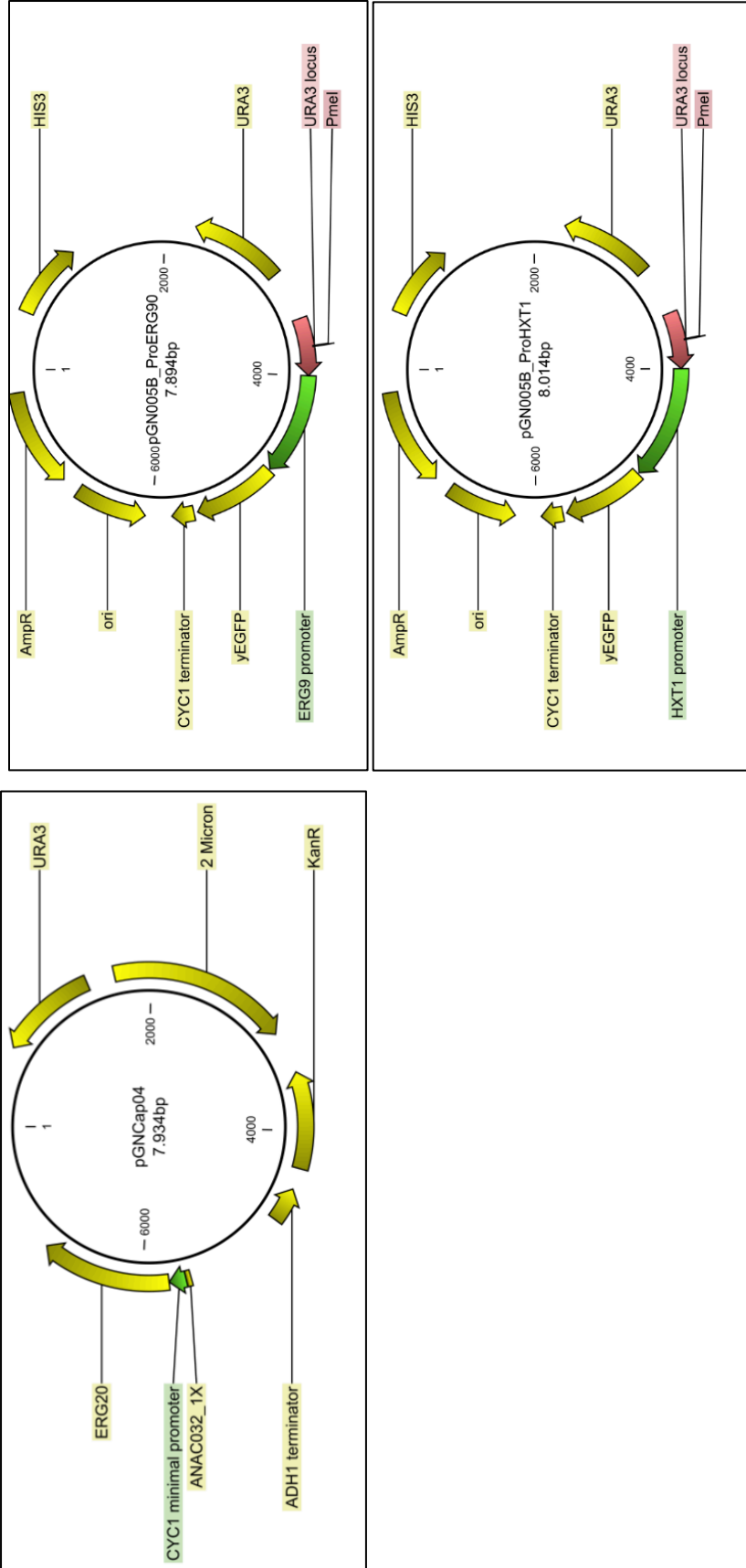
9. Partow, S., Siewers, V., Bjorn, S., Nielsen, J. and Maury, J. (2010) Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast*, 27, 955-964.
10. Borodina, I. and Nielsen, J. (2014) Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnol J*, 9, 609-620.
11. Westfall, P.J., Pitera, D.J., Lenihan, J.R., Eng, D., Woolard, F.X., Regentin, R., Horning, T., Tsuruta, H., Melis, D.J., Owens, A. et al. (2012) Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc Natl Acad Sci U S A*, 109, E111-118.
12. Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D. et al. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496, 528-532.
13. Mclsaac, R.S., Oakes, B.L., Wang, X., Dummit, K.A., Botstein, D. and Noyes, M.B. (2013) Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res*, 41, e57.
14. Brophy, J.A. and Voigt, C.A. (2014) Principles of genetic circuit design. *Nat Methods*, 11, 508-520.
15. Reider Apel, A., d'Espaux, L., Wehrs, M., Sachs, D., Li, R.A., Tong, G.J., Garber, M., Nnadi, O., Zhuang, W., Hillson, N.J. et al. (2017) A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, 45, 496-508.
16. Naseri, G., Balazadeh, S., Machens, F., Kamranfar, I., Messerschmidt, K. and Mueller-Roeber, B. (2017) Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth Biol*, 6, 1742-1756.
17. Kabadi, A.M. and Gersbach, C.A. (2014) Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression. *Methods*, 69, 188-197.
18. Lopez, J., Essus, K., Kim, I.K., Pereira, R., Herzog, J., Siewers, V., Nielsen, J. and Agosin, E. (2015) Production of beta-ionone by combined expression of carotenogenic and plant CCD1 genes in *Saccharomyces cerevisiae*. *Microb Cell Fact*, 14, 84.
19. Brown, S., Clastre, M., Courdavault, V. and O'Connor, S.E. (2015) *De novo* production of the plant-derived alkaloid strictosidine in yeast. *Proc Natl Acad Sci U S A*, 112, 3205-3210.
20. Walter, M.H. and Strack, D. (2011) Carotenoids and their cleavage products: biosynthesis and functions. *Nat Prod Rep*, 28, 663-692.
21. Verwaal, R., Wang, J., Meijnen, J.P., Visser, H., Sandmann, G., van den Berg, J.A. and van Ooyen, A.J. (2007) High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive transformation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. *Appl Environ Microbiol*, 73, 4342-4350.
22. Ohto, C., Muramatsu, M., Obata, S., Sakuradani, E. and Shimizu, S. (2009) Overexpression of the gene encoding HMG-CoA reductase in *Saccharomyces cerevisiae* for production of prenyl alcohols. *Appl Microbiol Biotechnol*, 82, 837-845.

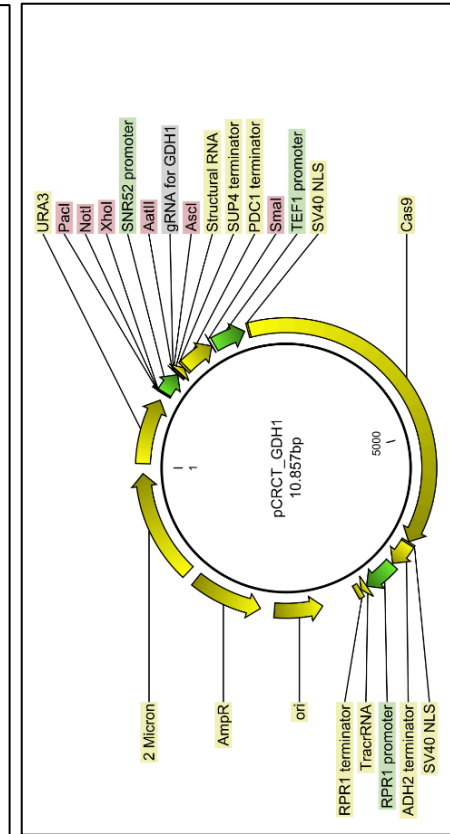
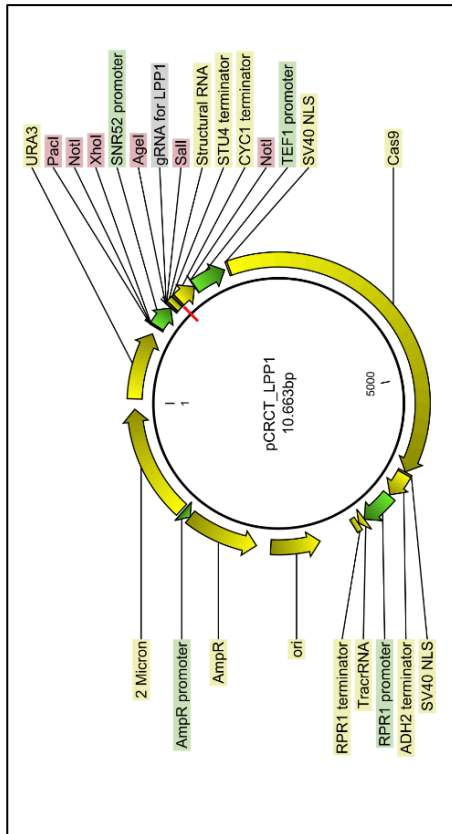
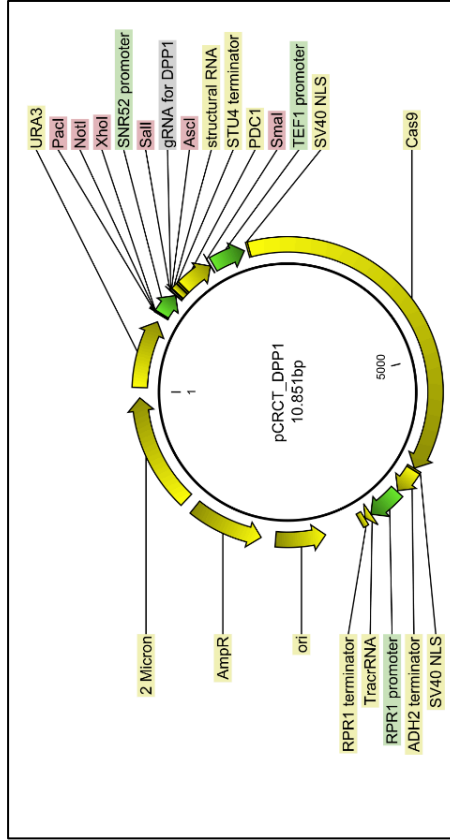
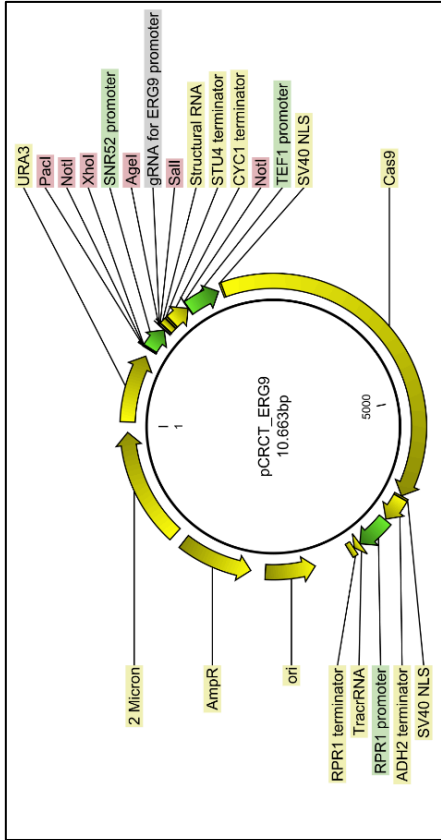
23. Kampranis, S.C. and Makris, A.M. (2012) Developing a yeast cell factory for the production of terpenoids. *Comput Struct Biotechnol J*, 3, e201210006.
24. Zhang, Y., Werling, U. and Edelman, W. (2012) SLICE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res*, 40, e55.
25. Messerschmidt, K., Hochrein, L., Dehm, D., Schulz, K. and Mueller-Roeber, B. (2016) Characterizing seamless ligation cloning extract for synthetic biological applications. *Anal Biochem*, 509, 24-32.
26. Gietz, R.D. and Schiestl, R.H. (2007) Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*, 2, 1-4.
27. Noskov, V.N., Chuang, R.-Y., Gibson, D.G., Leem, S.-H., Larionov, V. and Kouprina, N. (2010) Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. *Nat. Protocols*, 6, 89-96.
28. Hochrein, L., Machens, F., Gremmels, J., Schulz, K., Messerschmidt, K. and Mueller-Roeber, B. (2017) Assemblix: a user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Res*.
29. Jakociunas, T., Rajkumar, A.S., Zhang, J., Arsovska, D., Rodriguez, A., Jendresen, C.B., Skjodt, 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 Synth Biol*, 4, 1226-1234.
30. Ronda, C., Maury, J., Jakociunas, T., Jacobsen, S.A., Germann, S.M., Harrison, S.J., Borodina, I., Keasling, J.D., Jensen, M.K. and Nielsen, A.T. (2015) CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microb Cell Fact*, 14, 97.
31. G. An, D.B. Schuman and E.A. Johnson. (1989) Isolation of *Phaffia rhodozyma* mutants with increased astaxanthin content. *Appl Environ Microbiol* 55, 116–124
32. Shao, Z., Zhao, H. and Zhao, H. (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res*, 37, e16.
33. 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.
34. Breslow, D.K., Cameron, D.M., Collins, S.R., Schuldiner, M., Stewart-Ornstein, J., Newman, H.W., Braun, S., Madhani, H.D., Krogan, N.J. and Weissman, J.S. (2008) A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods*, 5, 711-718.
35. Lee, M.E., DeLoache, W.C., Cervantes, B. and Dueber, J.E. (2015) A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth Biol*, 4, 975-986.
36. Scalcinati, G., Knuf, C., Partow, S., Chen, Y., Maury, J., Schalk, M., Daviet, L., Nielsen, J. and Siewers, V. (2012) Dynamic control of gene expression in *Saccharomyces cerevisiae* engineered for the production of plant sesquiterpene alpha-santalene in a fed-batch mode. *Metab Eng*, 14, 91-103.
37. Nielsen, J. and Keasling, J.D. (2016) Engineering Cellular Metabolism. *Cell*, 164, 1185-1197.
38. Redden, H., and Alper, H. S. (2015) The development and characterization of synthetic minimal yeast promoters, *Nat Commun* 6, 7810.

39. Ghosh, A., Ando, D., Gin, J., Runguphan, W., Denby, C., Wang, G., Baidoo, E.E., Shymansky, C., Keasling, J.D. and Garcia Martin, H. (2016)  $^{13}\text{C}$  metabolic flux analysis for systematic metabolic engineering of *S. cerevisiae* for overproduction of fatty acids. *Front Bioeng Biotechnol*, 4, 76.
40. Shimizu-Sato, S., Huq, E., Tepperman, J.M. and Quail, P.H. (2002) A light-switchable gene promoter system. *Nat Biotechnol*, 20, 1041-1044.

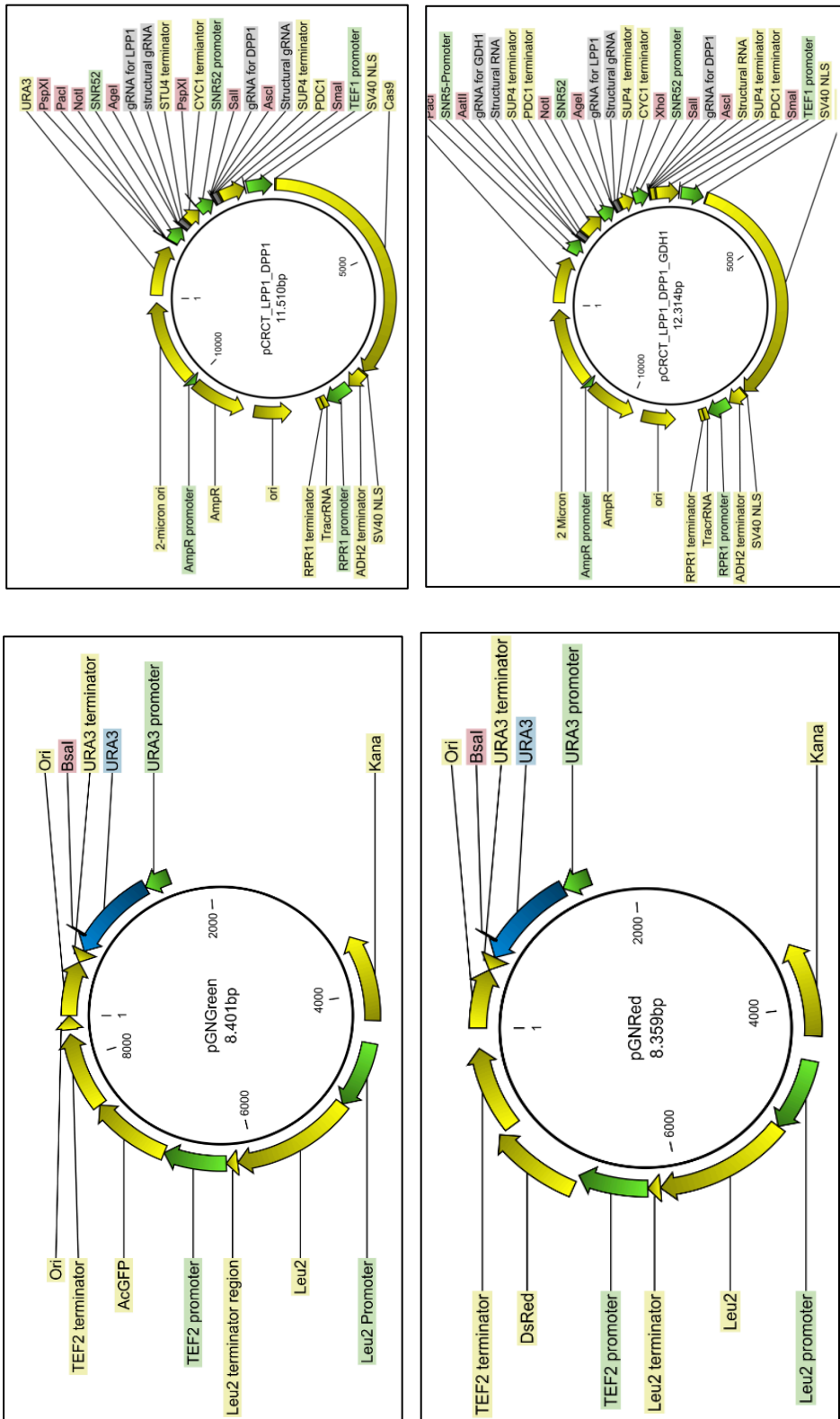
Supplementary Figure S1.



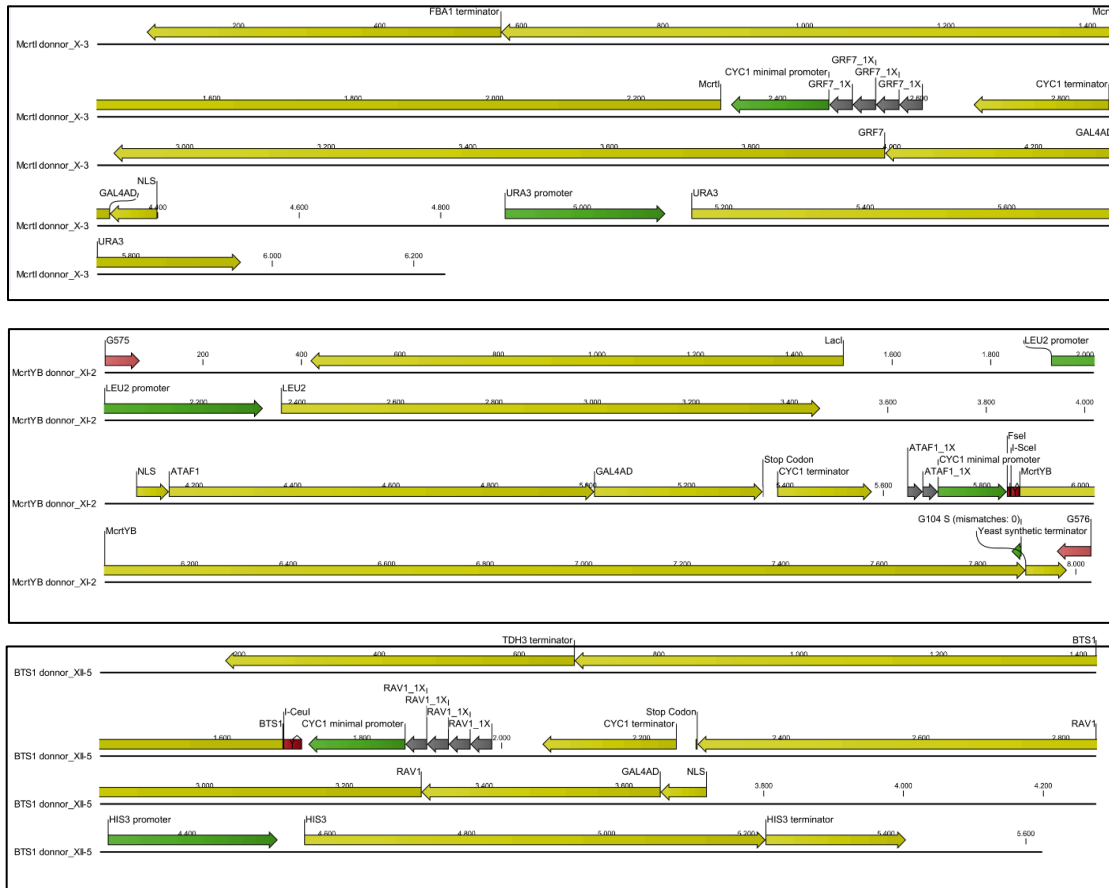




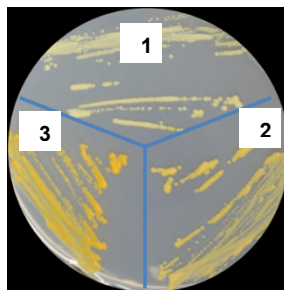




**Supplementary Figure S1. Plasmid maps.** The maps of plasmids used for CaPRedit-mediated methods are given.



**Supplementary Figure S2. Map of donors derive  $\beta$ -carotene production.** Donor contains an inducible plant-derived ATFP (the IPTG inducible *GAL1* promoter fused to a plant-derived ATF and its downstream terminator), the minimal *CYC1* promoter containing plant TF binding site, and the CDSs of  $\beta$ -carotene pathway fused to a yeast terminator. Donor sequence is flanked with 50-bp homology arms to up- and downstream of integration locus. Scheme showing Mrc1l-, BTS1- and Mrc1YB-CDS donors are shown in **Figure 2**.



**Supplementary Figure S3. Colors of  $\beta$ -carotene producing yeast strains.** Single colonies containing integrated  $\beta$ -carotene pathway CDSs (under the control of the ATF/BSs) for strains IMX672.1 (1), Gen 0.2 (2), and CaPRedit\_FPP 1.0 (3) were selected from the SC-Ura/-Leu/-His plates (the *URA3*, *LEU2*, and *HIS3* genes are encoded on the library modules). They then were re-streaked on inducing YPDA plates (2% galactose (w/v) and 20  $\mu$ M IPTG). All resulting colonies were of uniform colour.

## SUPPLEMENTARY METHODS

### General

All primers and plasmid sequences are given in **Supplementary Table S1** and **Supplementary S2**, respectively.

### Construction of yeast strain

Gen 0.1: The *S. cerevisiae* strain Gen 0.1 is derived from SCIGS22. Strain SCIGS22a has a CEN.PK background that is optimized for the FPP overexpression (1,2) and is auxotrophic for *URA3*. To compare the productivity of  $\beta$ -carotene in strains CaPRedit\_FPP 0.1 and SCIGS22a, we need positive selection scheme that involves auxotrophic marker genes. Therefore, SCIGS22a was object of further modification through implementing the Homology-Integrated CRISPR/Cas (HI-CRISPR) system for one-step multiple auxotrophic gene disruption (3). The HI-CRISPR strategy allows modification of multigenes in one step in yeast and uses plasmid pCRCT (Addgene, #60621) harbouring iCas9, a variant of WT Cas9 that increases the gene disruption efficiency, trans-encoded RNA (tracrRNA), and a homology-integrated crRNA cassette. The 100-bp dsDNA mutagenizing homologous recombination donor containing the gRNA sequence is inserted between two direct repeats for each target gene. Multiple donor and corresponding guide sequences can be integrated in pCRCT. The gRNA targeting 20-bp unique target sequences, to disruption each single gene, was selected via BLAST searches against the *S. cerevisiae* S288c genome (NCBI Taxonomy ID: 559292) to minimize off-target effects. Next, plasmid pCOMA and pCOMB were ordered as Gene synthesis from MWG to PCR-amplify fragments A (primers COMAfor/COMArev, on pCOMA) and B (primers COMBfor/COMBrev, on pCOMB). They were independently assembled into Bsal-digested pCRCT plasmid to generate plasmids pCOMC and pCOMD. Next, PCR- amplified fragment (primers COMDfor/COMDrev, on pCOMD) was assembled into Bsal-digested pCOMC to construct plasmid pCOM001. Plasmid pCOM001 contains gRNA targeting *leu2.a* (519 bp downstream of the *LEU2* start codon), *his3.a* (265-bp downstream of the *HIS3* start codon), *lys2.a* (799-bp downstream of the *LYS2* start codon), and *trp1.a* (245-bp downstream of the *TRP1* start codon). The plasmid was transformed into strain SCIGS22a cells to achieve frame-shift mutations for all four genes. The transformation was inoculated in liquid SC-Ura for overnight. Next, two-hundred microliters of a 10000-fold diluted cell culture were plated on SC-Ura plates. After 2 days, 50 colonies were randomly were streaked out onto four different selective plates (i.e., SC-Leu, SC-His, SC-Lys, SC-Trp). The colonies that did not grow on either of the four selective plates were streaked out on non-selective YPDA plate to remove the pCOM001. After four rounds of re-streaking single

colonies, in parallel, onto new YPDA agar medium and SC-Ura, we found a colony that was not able to grow on SC-Ura medium and therefore, the corresponding colony on YPDA plate was named strain Gen 0.1.

Gen 0.2: Strain Gen 0.1 was co transformed with PCR-amplified  $Pro_{TEF1}$ -Cas9-Ter<sub>CYC1</sub> fragment (GCASfor/GCASrev, on p414-TEF1p-Cas9-CYC1t, Addgene, #43802) as a donor, pCRCT (Addgene, #60621) encoding iCas9, and p426-SNR52p-gRNA.CAN1.Y-SUP4t (Addgene #47754) encoding gRNA targeting CAN1.w locus (4) The transformed cells were plated on SC-Ura/-Trp, as URA3 and TRP1 are encoded on pCRCT and p426-SNR52p-gRNA.CAN1.Y-SUP4t, respectively. Thereby, left and 50-bp right homology arms of Cas9-encoding donor lead to its integration into the CAN1.w locus of the genome. The successful integration was confirmed by colony PCR followed by sequencing). The positive colony was streaked out on non-selective YPDA plate to remove the plasmids. After several rounds of re-streaking onto new YPDA agar medium and SC-Ura/-Trp, a colony that was not able to grow on SC-Ura/-Trp was founded. The corresponding colony on YPDA plate was named strain Gen 0.2.

IMX672.1: To achieve the *lys2.a* frame-shift mutation in the IMX672 background, plasmid pCOM001 was transformed into the strain, as explained in Gen 0.1 construction section. Thirty colonies were streaked out on SC-Lys and YPDA agars. The colonies that did not grow on selective medium were streaked out on non-selective YPDA medium to eliminate the pCOM001. After several rounds of re-streaking colonies on YPDA and SC-Ura plates, we found a colony that was not able to grow on SC-Ura plate. The corresponding colony on YPDA plate was named strain IMX672.1.

#### **pCOM004**

The *LYS2* encoding fargment (LYSA\_for/LYSA\_rev, on pYC6Lys-TRP1URA3, Addgene #11010) and iCas9 encoding fragment (CASA\_for/CASA\_rev, on pCRCT (3)) were assembled into plasmid pTAJAK-92 (5) digested by NcoI and SphI.

#### **Entry vector X**

PmeI- digested pGN003B (6) was with treated with T4 DNA polymerase to remove 3' overhang ends and religated to generate pCOMPASS01. Moreover, PCR-amplified  $Pro_{TEF1}$ -*LacI*-Ter<sub>ADH1</sub> (primers LACfor/LACrev, on pCOMPASS01) was assembled into *AscI*/*FseI*-digested pLOA\_0\_1 (7) to generate pCOMPASS02. Next, PCR-amplified  $Pro_{TRP1}$  (primers PTRPfor/PTRPprev, on BY4741 genomic DNA) was assembled into the *NotI*/*PacI*-digested plasmid pCOMPASS02. Through this, *Bam*HI and *Sal*I sites were inserted downstream of  $Pro_{TRP1}$ . Moreover, a *PacI* site was inserted upstream of  $Pro_{TRP1}$ . The resulting plasmid was named pCOMPASS03. SbfI-

digested pCOMPASS03 was treated with T4 DNA polymerase to remove 3' overhang ends and re-ligated to generate pCOMPASS04. The pCOMPASS04 was digested with BamHI and Sall, and used in an assembly reaction with annealed single-stranded oligonucleotides CYCMfor/CYCMrev. Thereby, *MCS* and *X0*, the last 30 bp of *Pro<sub>CYC1mini</sub>*, were introduced between *Pro<sub>TRP1</sub>* and *E.coli<sub>ori</sub>*. The resulting plasmid was named Entry vector X.

### Acceptor vector A

NotI-digested pL0A\_0\_1 (7) was cut with *Pro<sub>KanaR</sub>* (primers PKANA\_for/PKANA\_rev, on pCR4-topo, Invitrogen), *Amp<sup>R</sup>* (primers AMP\_for/AMP\_rev, on pGN005B(5)) and *Ter<sub>KanaR</sub>* (primers TKANA\_for/TKANA\_rev, on pCR4-topo, Invitrogen) to generate pCOMPASS06. Next, pCOMPASS04 was digested with BamHI and NotI and used in three-way assembly reaction using PCR-amplified DNA parts *HIS3* and *Ter<sub>HIS3</sub>* (primers HISTERA\_for/HISTERA\_rev, on pGN005B (5)), *Pro<sub>HIS3</sub>* (primer PHIS\_for/PHIS\_rev, on pGN005B (6)), *Amp<sup>R</sup>* and *Ter<sub>KanaR</sub>* (primers AMPTER\_for/AMPTER\_rev, on pCOMPASS06). The generated plasmid was named Acceptor vector A.

### Acceptor vector B

BamHI/NotI-digested pCOMPASS04 was used in an assembly reaction using *LEU2* and *Ter<sub>LEU2</sub>* (primers LEUTER\_for/LEUTER\_rev, on pGAD424, TAKARA Bio, GenBank #U07647), *Pro<sub>LEU2</sub>* (primer PLEU\_for/PLEU\_rev, on pGAD424, TAKARA Bio, GenBank #U07647), *Cm<sup>R</sup>* and *Ter<sub>CmR</sub>* (primers CMRTER\_for/CMRTER\_rev, on pLD\_3\_4 (7)). The generated plasmid was called Acceptor vector B.

### Destination vector I

BamHI/NotI-digested pCOMPASS04 was used in an assembly reaction using *TRP1* and *Ter<sub>TRP1</sub>* (primer TRPTERC\_for/TRPTERC\_rev, on pGN003B(5)), *Pro<sub>TRP1</sub>* (primers PTRP\_for/PTRP\_rev, on pGN003B (5)), *TCS<sup>R</sup>* and *Ter<sub>TCSR</sub>* (primers TCSRTER\_for/TCSRTER\_rev, on pF2, Addgene #42520). The generated plasmid was called Acceptor vector C. Furthermore, BamHI/NotI-digested pCOMPASS04 was used in an assembly reaction using *LYS2* and *Ter<sub>LYS2</sub>* (primers LYSTERA\_for/LYSTERA\_rev, on pYC6Lys-TRP1URA3, Addgene #11010), *Pro<sub>LYS1</sub>* (primers PLYS\_for/PLYS\_ter, on pYC6Lys-TRP1URA3, Addgene #11010), *Gen<sup>R</sup>* and *Ter<sub>GenR</sub>* (primers GENTER\_for/GENTER\_rev, on pDEST32, ProQuest Two-Hybrid System with Gateway). The generated plasmid was named Acceptor vector D. Next, to remove the *URA3* encoding sequence (placed downstream of *E. coli<sub>ori</sub>* in pCOMPASS03), it was digested with PacI and ClaI. The ~5-kb vector was used in an assembly reaction using PCR-amplified fragment (primers

MURA\_for/MURA\_rev, on pCOMPASS03) to generate pCOMPASS07. Thereafter, NotI/PacI-digested pCOMPASS07 was used in an assembly reaction using PCR-amplified *Pro<sub>URA3</sub>* (primers URATERI\_for/URATERI\_rev, on pCOMPASS04) and PCR amplified *URA3 - Ter<sub>URA3</sub>* (primer PURAI\_for/PURAI\_rev, on pCOMPASS04) to generate pCOMPASS08. Through this, PacI and BamHI were inserted downstream (right) of *X0* site and NotI was inserted between *Pro<sub>URA3</sub>* and *URA3 CDS*. The pCOMPASS08 was then digested with BamHI and PacI and was used in an assembly reaction using *Spect<sup>R</sup> - Ter<sub>Spect<sup>R</sup></sub>* (primers SPECTERI\_for/SPECTTERI\_rev, on pCR8/GW/TOPO, TOPO Cloning Kit, TAKARA). Through this, the first seven nucleotides of *Spect<sup>R</sup>* were removed and PacI site was introduced downstream (left) of *Ter<sub>Spect<sup>R</sup></sub>*. The generated plasmid was named pCOMPASS09. Additionally, BamHI/SalI-digested pCOMPASS04 was used to assemble PCR-amplified *LYS2* and *Ter<sub>LYS2</sub>* (primers LYSTERI\_for/LYSTERI\_rev, on pYC6Lys-TRP1URA3; Addgene #11010), and annealed single-stranded oligonucleotides LYSX0\_for/LYSX0\_rev to result in pCOMPASS10. Moreover, BamHI/SalI-digested pCOMPASS04 was used to assemble PCR-amplified *TRP1* and *Ter<sub>TRP1</sub>* (primer TRPTERI\_for/TRPTERI\_rev, on pGN003B (6)), and annealed single-stranded oligonucleotides TRPX0\_for/TRPX0\_rev. The resulting plasmid was named COMPASS11. *LEU2* and *Ter<sub>LEU2</sub>* (primers LEUTERI\_for/LEUTERI\_rev, on pGAD424, TAKARA Bio, GenBank #U07647), and annealed single-stranded oligonucleotides LEUX0\_for/LEUX0\_rev were assembled into BamHI/SalI-digested pCOMPASS04. The generated plasmid was called pCOMPASS12. The *HIS3* and *Ter<sub>HIS3</sub>* (primers HISTERI\_for/HISTERI\_rev, on pGN005B<sup>17</sup>) and annealed single-stranded oligonucleotides HISX0\_for/HISX0\_rev introducing homology region to result in COMPASS13 were assembled in BamHI/SalI-digested pCOMPASS04. In the next step, PCR-amplified fragment *Ter<sub>HIS3</sub> - HIS3 - X0 - I-Ceul - Y1* (primers HIS\_for/HIS\_rev, on COMPASS13) was assembled in PacI-digested pCOMPASS09. The resulting plasmid was called pCOMPASS14. Next, PCR-amplified *LEU2* and its downstream terminator (primers LEU\_for/LEU\_rev, on pCOMPASS12) was assembled in FseI/PacI-digested pCOMPASS14. The generated construct was named pCOMPASS15. PCR-amplified *TRP1* and its downstream terminator (primers TRP\_for/TRP\_rev, on pCOMPASS11) was assembled in AscI/PacI-digested COMPASS15 to construct pCOMPASS16. PCR-amplified *TRP1* and its downstream terminator, *LYS2* and its downstream terminator (primers TRPI\_for/LYSI\_rev, on pCOMPASS16) were assembled in PacI-digested pCOMPASS15. The generated plasmid was called Destination vector I.

### Integration of a foreign $\beta$ -carotene pathway into the yeast genome

The coding sequences of NLS-GAL4AD-GRF7, NLS-ATAF1-GAL4AD, and NLS-GAL4AD-RAV1 were obtained by PCR using appropriate expression plasmids (6) as templates and the respective forward (ATF-for) and reverse (ATF-rev) primers. The corresponding binding sites (GRF7 4X, ATAF1 2X, and RAV1 4X) fused to the yeast minimal *CYC1* promoter were obtained by PCR using appropriate reporter plasmids (6) as templates and the respective forward (BS-for) and reverse (BS-rev) primers. They were assembled into the FseI/AsclI-digested Entry vector X. The resulted plasmids were named Entry vector X-GRF7-4X, Entry vector X-ATAF1-4X, and Entry vector X-RAV1-4X, respectively.

The promoters of the *E. coli* selection marker genes, yeast terminators, relevant parts of the CDSs of the  $\beta$ -carotene (**Supplementary Table S3**) biosynthesis pathway was amplified from different sources of genomic DNA or plasmids. The Entry vector X was digested with FseI and AsclI and three fragments including the CDS of the GOI, the yeast terminator and the promoter of the *E. coli* selection marker were inserted.

In the next step, PCR-amplified ATFP fragments (primers ATFPfor/ATFPprev, on Entry vector X-ATF/BS) together with the *Mcr1I*, *BTS1*, and *Mcr1YB* CDSs and their downstream terminators and the promoters of one *E. coli* selection marker genes (primers ENT\_for/MCRTI\_rev, on Entry vector X-Mcr1I, primers ENT\_for/BTS\_rev, on Entry vector X-BTS1, primers ENT\_for/MCR1YB\_rev, on Entry vector X-Mcr1YB, respectively) were cloned in Sall/EcoRI Destination vector I, FseI/AsclI Acceptor vectors A and B, respectively, carrying the CDS and terminator of corresponding *E. coli* selection markers (I: Spect<sup>R</sup>, A: Amp<sup>R</sup>, B: Cm<sup>R</sup>).

## REFERENCES

1. Lopez, J., Essus, K., Kim, I.K., Pereira, R., Herzog, J., Siewers, V., Nielsen, J. and Agosin, E. (2015) Production of beta-ionone by combined expression of carotenogenic and plant CCD1 genes in *Saccharomyces cerevisiae*. *Microb Cell Fact*, **14**, 84.
2. Scalcinati, G., Partow, S., Siewers, V., Schalk, M., Daviet, L. and Nielsen, J. (2012) Combined metabolic engineering of precursor and co-factor supply to increase alpha-santalene production by *Saccharomyces cerevisiae*. *Microb Cell Fact*, **11**, 117.
3. 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 Synth Biol*, **4**, 585-594.
4. 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 Synth Biol*, **4**, 585-594.
5. Ronda, C., Maury, J., Jakociunas, T., Jacobsen, S.A., Germann, S.M., Harrison, S.J., Borodina, I., Keasling, J.D., Jensen, M.K. and Nielsen, A.T. (2015) CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. *Microb Cell Fact*, **14**, 97.
6. Naseri, G., Balazadeh, S., Machens, F., Kamranfar, I., Messerschmidt, K. and Mueller-Roeber, B. (2017) Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth Bio*, **6**, 1742-1756.



7. Hochrein, L., Machens, F., Gremmels, J., Schulz, K., Messerschmidt, K. and Mueller-Roeber, B. (2017) AssemblX: a user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Res*, **45**, e80.

## 7 DISCUSSION

The yeast *Saccharomyces cerevisiae* is a widely used chassis for various applications in synbio. Nowadays, advanced cloning methods as well as cheaper and faster gene synthesis are making it easier to implement complex genetic designs in yeast. However, the goals of synbio projects are getting more and more technically and biologically ambitious, and the tools need to keep up developing and improving.

For example, the enzymes for the biosynthesis of complex products such as flavonoids, glucosinolates, and steroidal glycoalkaloids in plants are encoded by more than 10 genes whose co-expression is tightly regulated.<sup>87, 118</sup> In such pathways, transcription of each gene is controlled by its own promoter and is coordinated by TFs.<sup>76</sup> Once pathways of complex secondary metabolites are introduced in yeast, synchronizing expression, which is often critical for efficient metabolite production, can present a challenge.<sup>119</sup> This results in a desire for tools allowing orthogonal control of heterologous gene expression.<sup>44</sup> In fact, for the regulation of individual gene expression from a complex system, different capacities of transcriptional activation are needed. Because the transcriptional machinery of yeast, like other eukaryotic organisms, is not able to support a polycistronic gene expression, yeast employs individual promoters to control the expression of each single gene.<sup>39</sup> Synthetic biologists tried to translate the operon-based concept of prokaryotic organisms to *S. cerevisiae* through transferring the 5'UTRs as linkers between coding sequences, although they have not achieved so much success.<sup>120-122</sup> In contrast, they were successful in developing promoter collections and ATFPs that can provide valuable tools to facilitate the build phase.<sup>17, 42, 44, 46</sup>

To generate robust and orthogonal regulation of gene expression control, TALE and CRISPR / dCas9 systems are commonly used.<sup>42</sup> However, the repeat structure and quite large size of TALE-derived ATFP<sup>38, 42, 46</sup> and the inferior degree of transcriptional activation by CRISPR-derived ATFP<sup>42, 44</sup> and their promoter pairs make them undesirable for synthetic biology purposes. Hence, most metabolic engineering projects in yeast rely on a small set of yeast constitutive and inducible promoters.<sup>24, 41, 92</sup> In particular, synthetic biology projects tend to repetitively implement strong promoters of yeast genes such as the *GAL1* or *TDH3*.<sup>8,24, 123</sup> However, the repeated use of identical promoters may result in construct instability, as yeast has a quite active homologous recombination machinery.<sup>73</sup> Therefore, in this study, a new class of

## DISCUSSION

---

ATFs was generated based on heterologous TFs, such as those derived from plants (Chapter 4).

The plant TFs are supposed to be proper candidates to generate orthogonal ATFP, because plants are evolutionary far from yeasts and thereby the risk of interactions between plant TFs and yeast regulatory elements is low. Therefore, plant-derived ATFPs less likely interfere with yeast endogenous regulatory networks. Moreover, higher plants have more than 2,000 TFs,<sup>124-125</sup> posing chances for building orthogonal ATFs in yeast. Although plant and animal TFs are consistently expressed in yeast for one-hybrid and two-hybrid analyses, they have so far not been used to establish orthogonal regulatory systems.<sup>17</sup> In this study, several plant TF families, including the NAC TFs, were tested to control gene expression in yeast. The results (published in ACS Synthetic Biology journal)<sup>17</sup> showed such TFs are in principle suitable for establishing orthogonal gene regulatory networks in yeast and other heterologous systems.

The drawback with regulators derived from heterologous TFs, as opposed to CRISPR-derived ATFP, is that they cannot be easily designed to target specific binding sites, but rather must be paired with their endogenous binding sites.<sup>125</sup> Therefore, first a driver / reporter platform was established to assess the activation strength and specificity of heterologous TF-binding motifs in yeast. This platform allows selection of plant-derived binding sites that act orthogonally to the endogenous gene regulatory networks of yeast. Specifically, candidates from 14 different TF families of *A. thaliana* that do not exist in the yeast *S. cerevisiae*, were tested. Relatively high basal fluorescence output was observed for the binding sites of three TFs (ANAC032,<sup>126</sup> DREB2A,<sup>35</sup> and LFY,<sup>127</sup>) representing interaction between these binding sites and endogenous yeast TFs. These TFs were therefore excluded from further studies due to insufficient specificity. For plant-derived ATFPs generated using DOF1<sup>128</sup>-, MYB61<sup>129</sup>- and WRKY6<sup>130</sup> and GAL4 AD only low transcriptional output was observed, and they were therefore also excluded. The platform successfully identified nine suitable plant TFs with orthogonal binding sites including ANT,<sup>131</sup> RAV1,<sup>132</sup> WRKY6,<sup>130</sup> GRF7<sup>133</sup> and GRF9,<sup>134</sup> JUB1,<sup>135</sup> ORE1,<sup>136</sup> ATAF1,<sup>137</sup> and ANAC102.<sup>138</sup> Transcriptional output of these driver / reporter combinations varied over a wide range, reaching levels as high as 2,000-fold induction of gene expression. Some of the synthetic transcriptional units triggered expression output which was 6- to 10-fold higher than that of the strong yeast *TDH3* promoter.

## DISCUSSION

---

Low to very high transcriptional output observed for ATFP constructed using different families of plant TF might probably be due to several reasons, in particular, differences in the binding strength of TFs to their targets sites that is affected by a number of factors in addition to binding specificity, including interactions with other transcription regulatory proteins and subcellular trafficking components. Moreover, eukaryotic TFs may contain a repressor domain that influences the activity of the activation domain fused to it and the transcriptional output of ATFP derived from different plant TFs. Most of the chosen plant-derived ATFPs (except ANT and RAV1) with significant transcriptional output contain a native AD at the C-terminus. For ATFs obtained by adding an extra AD (*i.e.*, the GAL4 AD) to the N-terminus of plant TFs (AD-TF-AD), transcriptional output mostly increased with the number of binding sites. This association between transcriptional output and binding site copy number was less apparent for ATFs (except ANT and RAV1) carrying the GAL4 AD at the C-terminus (TF-AD-AD). One possible explanation is that greater flexibility in the arrangement of the plant and artificial ADs around the DNA binding domain facilitates interactions with the yeast general transcription activation machinery. This result suggests that keeping a larger distance between two different ADs (the native plant AD and the yeast AD) in the primary sequence of the ATFs usually leads to a transcriptional output of plant-derived ATFP this is positively correlated to the number of the binding site. In other words, effects arising from the ATF structure is reduced, which is be wanted, when synthetic biologists need to predict the output of a designed system. As mentioned above, the ATFPs derived from ANT and RAV1 are the exceptions to this trend. Previously, it has been reported that ANT contains two ADs and RAV1 contains two DBDs in the N- and C-terminus.<sup>132</sup><sup>139</sup> Such architectural characteristics of ANT- and RAV1-derived ATFs may be causing them to behave differently in the platform.

In regards to the fact that the position of the TF relative to the GAL4 AD (N- versus C-terminal) can have a strong effect on the binding and/or transactivation capacity of plant-derived ATFPs, NLS-GAL4AD-ANAC102 drove high fluorescence output, but no fluorescence output was detected with NLS-ANAC102-GAL4AD. Previously, it has been shown that GFP-ANAC102 has nucleus localization while ANAC102-GFP is localized in the chloroplast of plant.<sup>140</sup> Therefore, C-terminal ADs might interfere with nuclear trafficking. Another surprising observation was that both ORE1-derived TFs, *i.e.* NLS-GAL4AD-ORE1 and NLS-ORE1-GAL4AD, showed a similar transactivation capacity in both, IPTG-induced and non-induced cells. Previously, some NAC TFs were characterized for protein intrinsic disorder (ID) that causes tertiary structure not to be fixed.<sup>141</sup> Therefore, ORE1 might be able to strongly interact with yeast proteins, while the low

## DISCUSSION

---

amount of ORE1-derived ATF is present in the yeast cell (due to its leaky expression in non-inducing medium).

The collection of ATFs described in the manuscript will be of immediate relevance to researchers working with yeast as a chassis for synthetic biology and biotechnology applications. Further, the ATFP screening platform allowed us to extract three simple guidelines for non-expert users to construct their own heterologous-based ATFPs:

- (i) If very high transcriptional output is required, TF-AD-AD with two binding sites would be the preferred choice, especially when the ATF is derived from a plant NAC TF.
- (ii) If continuously increasing transcriptional output of one plant-derived ATFP is desired, AD-TF-AD with one, two and four binding sites can be selected to have weak, medium and quite high transcriptional output.
- (iii) The numerous NAC TFs encoded by plant genomes are excellent candidates for the establishment of orthologous transcription regulatory circuits in synthetic biology projects due to their strong transactivation capacity in yeast and the virtual absence of their binding sites in the yeast genome.

Furthermore, the insights gained from the use of plant TFs in yeast may also trigger researchers working with different hosts to implement plant-derived ATFPs in their respective systems as well as implementing TFs derived from other organisms to build ATFPs for synbio purposes in yeast. Thus, our findings will be of interest to a broad range of synbio users.

Furthermore, the results showed that plant TFs can be used to build regulatory systems encoded by centromeric and episomal plasmids. Green and red fluorescence proteins were used to monitor the transcriptional outputs of two plant-derived ATFPs encoded on the same plasmid. In the case of the plasmid, two individual (IPTG- and ATc-) inducible *GAL1* promoters could independently activate transcription of two plant-derived ATFPs. Such transcription regulatory modules that can function independently of each other are needed for full exploitation of genetic circuit construction tools for synbio projects.<sup>3, 25, 142</sup> Later, ATFPs were implemented to express the genes of a desired biosynthetic metabolic pathway in yeast.

One major goal in synbio projects is implementing hosts, such as yeast, as a cell factory to enhance the production of new products.<sup>143</sup> Biosynthetic pathways typically consist of multiple genes whose individual protein products convert an initial substrate, through some number of intermediate steps, into a desired end product.<sup>143</sup> Although engineered organisms are recently being used in closed industrial scale, the amount of products are typically in the range of  $\mu\text{g/l}$

## DISCUSSION

---

to mg/l.<sup>144</sup> Until now, production of only a few compounds, such as 1,3-propanediol,<sup>145</sup> 1–4-butanediol<sup>146</sup> and artemisinin<sup>8</sup> has reached commercial scale. Such a commercialization requires detailed knowledge of microbial physiology, stress response, and metabolism with especial attention to the carbon and energy balance. Furthermore, the environment in which the cells operate will have a significant impact on the functionality of engineered cells.<sup>65</sup> Hence, the tools, modules, circuits and systems that work well in a laboratory environment, do not work as well in much more complex natural environments.<sup>94</sup> Synthetic biologists require slowly evolving and adjusting the engineered system to such environmental and contextual differences. A more sophisticated solution might be to precisely consider the interaction of the exogenous pathway with host's endogenous metabolism.<sup>147</sup> To reach a wanted biochemical, physiological or developmental output in a fully functional synbio setting, computational modelling of the interaction and co-action of diverse biological parts would be the final aim. Currently, however, such an aim is in far reach for any, even very well characterized organisms like *S. cerevisiae*, as a major task thereby concerns establishing appropriate expression levels for all genes in the pathway to achieve maximal product output.<sup>148</sup> Prominently, *a priori* knowledge about the optimal expression levels of all genes in a synthetic gene regulatory network is not usually available.<sup>69</sup> Moreover, building multi-gene constructs that give an optimal output is very time consuming.<sup>69, 109</sup> Constructing all possible combinations of various promoters and genes of pathway becomes very laborious as the number of promoters or / and genes increase. For example, a five-gene pathway where each gene is expressed from only three different promoters already requires the building of  $3^5 = 243$  different constructs, while with ten genes and five promoters each already  $5^{10} = 3125$  constructs are theoretically possible.

Smartly designed combinatorial libraries can increase the number of library variants and are thus a greatly valuable tool to optimize and maximize yield and productivity for commercialization and scale-up production.<sup>110</sup> Despite of some progress in pathway optimization obtained in the last years,<sup>149-152</sup> MAGE (multiplex automated genome engineering) is still the most effective technology for combinatorial optimization of metabolic engineering projects.<sup>153</sup> MAGE implements short synthetic oligonucleotides containing ribosome binding site (RBS) flanked by HRs to simultaneously introduce genetic modifications at multiple loci of the *E. coli* genome.<sup>153</sup> However, the application of MAGE in *S. cerevisiae*, an ideal host for industrial production of compounds, was not successful, most likely because of poor efficiency in gene replacement mediated by short oligonucleotides in yeast.<sup>154</sup> Recently, Lian *et al.* developed a new combinatorial strategy based on an orthogonal tri-functional CRISPR system that combines

## DISCUSSION

---

transcriptional activation, transcriptional interference, and gene deletion (CRISPR-AID) for metabolic engineering purposes in *S. cerevisiae*.<sup>65</sup>

Here, the COMPASS (Chapter 5), COMbinatorial Pathway ASSEMBly method was established for rapid expression balancing of metabolic pathway genes in the yeast *S. cerevisiae* that accomplishes the requirements needed for such demanding projects.

COMPASS allows optimizing the expression of pathway genes towards optimal product output without *a priori* knowledge of the best combination of expression levels of the individual genes. The system employs ATFPs to control the expression of pathway genes over a wide range.<sup>17</sup> In its current realization, COMPASS allows the fast combinatorial assembly of up to ten pathway genes, each transcriptionally controlled by nine plant-derived ATFPs, covering a 10-fold difference in expression strength. COMPASS employs a recombinatorial cloning to generate all possible combinations between the ATFPs and the coding sequences (CDSs) of enzymes required to build the metabolic pathway, enabling a theoretical complexity of  $9^{10} \approx 3.5 \times 10^9$  different assemblies.

By coupling COMPASS with multi-locus CRISPR/Cas9-mediated genome modification, it is possible to generate libraries of stable yeast (*S. cerevisiae*) variants with a complexity of thousands to millions of different members through only four (combinatorial) cloning reactions followed by the decoupled integration of the constructs into the yeast genome. Of note, to achieve this very large number of constructs, we employed only nine inducible plant-derived ATFPs of our previously reported collection of 106 plant-derived regulators.

COMPASS is based on overlap-based cloning methods, like SLICE,<sup>156</sup> Gibson, and NEBuilder HiFi assembly<sup>157</sup> (*in vitro*), and TAR (*in vivo*).<sup>158-159</sup> Because some assembly products are not stable or clonable in *E. coli* due to, for example, high AT content and Z-DNA-like structures or outer membrane of *E. coli* is not easily permeable for macromolecules,<sup>160</sup> the COMPASS vectors are equipped with yeast selection markers allowing positive selection for TAR cloning in yeast. While NEBuilder HiFi is preferred for generating ATFP-CDS module, TAR is the desired approach for combining the multi ATFP-CDS modules of pathways.

Screening for correct gene assembly from combinatorial cloning is typically a tedious but critical step because assembled constructs often lack DNA elements or the DNA fragments are shuffled in an uncontrolled manner. Therefore, giving the possibility of positive selection after combinatorial gene assembly reduces the rate of fragment mis-assembly and allows sequence

## DISCUSSION

---

independent screening for successful assembled constructs.<sup>155</sup> The positive selection in COMPASS is provided with including non-functional gene of *E.coli* and yeast selection markers (due to lack of an essential sequence) in the cloning vectors, whereby the missing sequences of the selection markers are included within the insert. Hence, *in vitro* or *in vivo*-mediated successful assemblies will produce the functional selection markers. Consequently, it allows the possibility for rapid construction of combinatorial library of ATFP upstream of multigenes into multi vectors in a single cloning reaction tube (see Chapter 5).

The presence of the yeast and *E.coli* positive selection markers in the integrating modules provides a higher level of control for successful integration (a critical issue in combinatorial approaches). However, the number of suitable selection markers is limited, and therefore the current version of COMPASS can be used for the optimization of limited number of genes (pathways with up to ten genes; see Chapter 5). Given the current cyclical and iterative nature provides COMPASS with the optimization of greater than ten-gene pathways. One further work that needs to be done is flanking selection markers by *loxP* sites. Hence, ten marker genes of the COMPASS vectors, each need to be flanked with *loxPsym* sites. The markers can be removed after successful integration of the first ten genes of interest using Cre recombinase allowing recombination at *loxPsym* sites.<sup>161</sup> In addition, CRISPR/Cas9-mediated genome modification can be utilized to remove or mutate the selection marker coding sequences,<sup>56</sup> and therefore pathways with more genes can be established.

For the heterologous engineering of a metabolic pathway in yeast, easy manipulation of plasmid-based system makes it favorable for synthetic biologists. In contrast, segregational and/or structural instability of plasmid-based pathway expression,<sup>30</sup> and due to the higher expression level achieved by integration of pathways into the chromosomal DNA<sup>17</sup>, COMPASS was designed to have the entire pathway (all CDSs) contained on the YAC from Level 2 that can also be integrated into the genome. Moreover, COMPASS implements a high-efficiency CRISPR/Cas9-mediated modification system to facilitate the integration of a library of multiple groups of donors at multiple loci, whereby each group is integrated into a single locus of different yeast cells.<sup>56</sup> Each group of donors contains a library of ATFP upstream of one CDS.

COMPASS cloning can be a relatively flexible and fast approach for optimization of any interested pathway expression, because each CDS, in COMPASS vectors, can be replaced with any other CDS of interest, because the CDSs are flanked by unique restriction enzyme cleavage sites. This would be well suited to other synbio goals such as building synthetic signal



## DISCUSSION

---

transduction pathways, complex proteins, and gene regulatory networks (GRN). Although the current version of COMPASS is equipped with the collection of plant-derived ATFPs to control the gene expression, other regulators such as TALE- and CRISPR/Cas9-derived ATFPs can be easily implemented as well. For example, Cas9-derived ATFs and collection of synthetic promoters containing one to multiple copies of its binding sites can be assembled in the Entry vector X in a combinatorial manner. This property will allow COMPASS to be adapted for the assembly of pathway libraries for the production in hosts other than yeast (in the cases where a plant is the desired host organism).

In general, the size of library is defined by the number of pathway genes and ATFPs used to regulate their expression (size of library = number of ATFPs<sup>number of genes</sup>). In the current version of COMPASS, nine out of 106 members of the plant-derived ATFP collection were utilized to express up to nine genes. We therefore are theoretically able to produce a combinatorial library with  $9^9$  (387 420 489) members. After successfully generating the pathway library using currently available combinatorial approaches, a big challenge remains in respect to the ability to screen the output to support the established combinatorial method in the lab and to find the best producing strain in industrial scale.

In the first attempt, the optimization ability of COMPASS was studied through constructing a library of colorful product, namely  $\beta$ -carotene.<sup>123</sup> The production of  $\beta$ -carotene was studied in a yeast strain with a CEN.PK background and an optimized strain which also has a CEN.PK background but has additionally been optimized for the production of farnesyl pyrophosphate (FPP), a precursor of  $\beta$ -carotene biosynthesis.<sup>81-82</sup> The top  $\beta$ -carotene producer (in the optimized background) yielded  $0.81 \pm 0.25$  mg  $\beta$ -carotene  $g^{-1}$  cdw. Moreover, the results strongly indicate that a combination of weak, medium, and strong ATFPs upstream of the pathway genes is required for high-level  $\beta$ -carotene production. We then established the pathways for the colorless chemical  $\beta$ -ionone, a downstream product of  $\beta$ -carotene, in the best producers. The enzyme RiCCD1 converts  $\beta$ -carotene to  $\beta$ -ionone leads to yeast cells that are less intensely colored than  $\beta$ -carotene-producing cells.<sup>82</sup> Medium or high expression of RiCCD1 (in 86% cases) was needed to produce more  $\beta$ -ionone in high  $\beta$ -carotene accumulators. Moreover, more  $\beta$ -ionone accumulated in the wild type than the optimized strain (for  $\beta$ -carotene production), demonstrating that superior  $\beta$ -carotene accumulation is not *per se* sufficient for a high-level accumulation of  $\beta$ -ionone in yeast.

## DISCUSSION

---

After proving the functionality of COMPASS as a novel combinatorial cloning method by assembling biochemical pathways for  $\beta$ -carotene and  $\beta$ -ionone production in *S. cerevisiae*, this then extended by establishing yeast cells jointly producing  $\beta$ -ionone and naringenin, whereby the accumulation of naringenin was evaluated at the single-cell level using a recently reported naringenin biosensor.<sup>113</sup> NG production is sensed by a prokaryotic transcriptional activator FdeR and its corresponding binding site (*FdeO*) located upstream of reporter gene in yeast, where the diversity of NG production is monitored via FACS. Analyzing less than 0.000005% of the theoretical complexity of the library revealed that approximately 30% of the library members show no or low level of both  $\beta$ -ionone and NG productions, emphasizing the importance of developing combinatorial optimization approaches. The better NG producers harbor medium and strong ATFP units upstream of NG genes, while better  $\beta$ -ionone producers harbor combination of weak, medium and strong ATFPs upstream of  $\beta$ -ionone genes. Taken together, one strain, called Narion 14, was identified as the best producer. In the collection, it produces the medium level of NG and the highest level of  $\beta$ -ionone ( $0.18 \pm 0.017 \text{ mg g}^{-1} \text{ cdw}$ ). In Narion 14, the expression of NG and  $\beta$ -ionone pathway genes is controlled by weak/medium and weak/medium/strong ATFP regulators, respectively. Here, the accumulation of NG was detected using a biosensor system sensitive to NG concentration. Coupling COMPASS with biosensor-based screening methods allowing its future implementation in other single cell-based screening procedures. Hopefully, once biosensors can be constructed for any desired product, high-throughput screening with tools such as FACS and microfluidics will be possible. For example, synthetic riboswitches that respond to different chemicals have been developed for sensing applications.<sup>114-115</sup>

The highlights of COMPASS toolkit are:

- (i) Establishment of a high-throughput cloning method for the combinatorial assembly of nine chemically inducible artificial transcription factors (ATFPs) with up to 10 different pathway genes;
- (ii) COMPASS allows building thousands of different plasmids for pathway engineering in a single cloning reaction tube;
- (iii) Identification of successful gene assemblies by positive selection for cell growth for both, *in vivo* and *in vitro* homologous recombination methodologies;
- (iv) Adapted for plasmid- and genomic integration-based metabolic engineering projects;
- (v) Easy to implement other classes of ATFs and other inducers, e.g. light;

## DISCUSSION

---

(vi) Easy to adopt to other biochemical metabolic engineering projects or other synthetic biology aims e.g. for building protein complexes, and gene regulatory networks in yeast *S. cerevisiae*.

In the present work COMPASS was used for the transcriptional optimization of multigene pathways. Gene shuffling is a simple method for the generation of sequence libraries using a family of related genes and is another interesting scope needs to be considered for pathway optimization.<sup>162</sup> In COMPASS, the impact of combinatorial gene shuffling was shown briefly; two different GGPPS, the key enzyme in MVA pathway catalyzing FPP to GGPP, namely GGPPSbc (from *T. baccata* x *T. cuspidate*)<sup>91</sup> and BTS1 (from *S. cerevisiae*)<sup>123</sup> were used for  $\beta$ -carotene pathway engineering. The result showed that BTS1 containing strains produced more  $\beta$ -ionone, downstream product of GGPP, than GGPPSbc containing strains. In addition to altering the ATFPs and the CDSs, induction optimization, other cheaper and reversible inducers such as light-inducible systems<sup>163-164</sup> can improve the production level of the chemicals. Furthermore, development and optimization of fermentation process for large-scale production<sup>14</sup> are further scopes that need to be considered.

A higher amount of  $\beta$ -carotene, than in the wild type, was observed in the strain that has the CEN.PK background<sup>82</sup> and is optimized for the production of FPP (key branch point for the biosynthesis of all isoprenoids). This confirms the metabolic engineering efforts, combined with the developing a yeast cell factory for the production of key precursor will enable the industrial and environmentally friendly production of a wide range of compounds. Therefore, in next effort the metabolic engineering of microbial strains to produce high levels of precursors of heterologous pathway was studied. The CaPRedit approach (Chapter 6) was developed for speeding up the strain modification procedure and large-scale changing of enzyme expression (high expression for key enzymes, no to low expression for unwanted enzymes) in desired stage of growth, through implementing CRISPR/Cas9-mediated one-step multigene modification system<sup>56</sup> and inducible plant-derived ATFPs (10-fold stronger than the yeast constitutive strong *TDH3* promoter),<sup>17</sup> respectively.

CRISPR/Cas9-mediated integration allows highly efficient integration of DNA donors without applying selection pressure. In CaPRedit, controllable plant-derived AFTP(s) are guided and replaced with the sequence(s) that need to be modified (replaced or deleted) via a gRNA and the endonuclease active Cas9 and mediates up- or downregulation of target genes. Generally, it is favorable if yeast grows enough and then produces the target products. Therefore, utilizing

## DISCUSSION

---

IPTG inducible promoter to control expression of strong plant-derived ATFs in CaPRedit makes it possible to express ATFPs only after the yeast culture has expanded sufficiently.

As an application for this approach, an efficient *S. cerevisiae* cell factory for isoprenoid production was constructed by redirecting carbon towards the central precursor to nearly all isoprenoid products, including squalene<sup>165</sup> and  $\beta$ -carotene.<sup>123</sup> To increase FPP production and minimize FPP consumption for the metabolites that compete with  $\beta$ -carotene synthesis from FPP three types of modifications need to be made<sup>81-82</sup>: (i) Overexpression of *GDH2*, *tHMG1*, and *ERG20* through integrating another copy of them under the control of inducible strong plant-derived ATFPs; (ii) Downregulation of *ERG9*, an enzyme which converts FPP to essential metabolites, through replacement of the native promoter of *ERG9* with a synthetic promoter containing the binding site of a plant TF (with transcriptional output less than *Pro<sub>ERG9</sub>*) that leads to constitutive and low expression of *ERG9*; and (iii) Deletion or inactivation of *DPP1*, *LPP1* and *GDH1* to minimize FPP consumption for the biosynthesis of unnecessary metabolites. Subsequently, the amount of  $\beta$ -carotene can be quantified in the modified strain to compare the efficiency of CaPRedit-optimized strains with the previously optimized strain (using yeast constitutive promoters).

In this study, it was demonstrated that upon only overexpression of key enzymes *GDH2*, *tHMG1*, and *ERG20* by three strong plant-derived ATFPs (2.5- to 5.5-fold stronger than yeast *TDH3* promoter), a *S. cerevisiae* strain producing  $0.61 \pm 0.044$  mg  $\beta$ -carotene/g (dw) was generated. In contrast, only  $0.45 \pm 0.05$  mg  $\beta$ -carotene/g (dw) was achieved for the previously reported optimized strain harbouring more modification to reduce FPP consumption by competing pathways, in addition to constitutively overexpressed *GDH2*, *tHMG1*, and *ERG20* genes. Moreover, the growth of cells without the induction of ATFPs was measured, under the statement that undesired interactions with the host genome could impose a fitness cost on the cells using previously reported flow cytometry-based high-throughput technique.<sup>166</sup> This approach is independent from the colony size or optical density measurements and relies on the fluorescent protein expression that allows monitoring slight to no effect of ATF presence on the yeast growth. The result demonstrates that implementing yeast constitutive promoter had major effect on yeast growth. On the contrary, little growth defect when plant-derived ATFPs were integrated in the genome (but not expressed) was observed. In fact, the inducible system provides the possibility to express the genes when yeast grows enough by means of adding inducer to the system. In this way, cells are not exposed the unnecessary stresses caused by redirecting many

## DISCUSSION

---

necessary cellular resources towards target protein (and mRNA) production. This outcome can be undesirable when cellular processes are needed for precursor, co-factor, or biomass production during the lag and log phases of the yeast growth. Thereby, in non-inducing medium, in the absence of plant-derived regulators, the yeast cells grow for maximal gain of biomass. This provided the yeast cell factory with maximal resources required for production of desired metabolite after inducing expression of the plant-derived regulator. The results provide additional evidence that reduced expression strength could lead to higher production through decreased metabolic burden. The generalized CaPRedit method presented above can be quickly and easily applied to a wide variety of systems and metabolites such as biofuels and pharmaceuticals to improve yields, and titers of interested product.

To control the complex interactions between metabolic pathways, researchers usually need to change production of various targets in different ways, for example increasing the expression of rate-limiting enzymes, decreasing the expression of essential genes, and eliminating the expression of competing pathways.<sup>61</sup> The CaPRedit strategy presented here could be applied for increasing any metabolite production in *S. cerevisiae* by increasing production of the endogenous rate-limiting or/and heterologous enzymes of a desired metabolic pathway and decreasing the expression of essential genes for the host metabolism. It was demonstrated that CaPRedit is a powerful tool because of implementing synchronized, inducible, and strong, plant-derived ATFPs that are 2.5- to 5.5-fold stronger than the yeast *Pro<sub>TDH3</sub>* promoter and is a fast approach due to using highly efficient CRISPR/Cas9 genome editing that allows one-step multiloci targeted genomic integration without selective pressure.

Prior to this work, strong ATFPs were lacking from the toolbox of synthetic biologists. Here, a new class of superior ATFPs, including plant-derived ATFPs up to 10 fold stronger than the *TDH3* promoter of yeast, was introduced.<sup>17</sup> Microbial production systems become more complex that means the high expression level of one enzyme might be desirable, while low expression of another enzyme is desirable. Therefore, the necessity for fine-tuning of expression systems increases. Hence, the collection of plant-derived ATFPs with wide range of transcriptional output can pave the way to achieve such a tuning of expression. More interestingly, the COMPASS method was developed for transcriptional optimization of a desired metabolic pathway using the collection of plant-derived ATFPs. The current version of the COMPASS approach aims to optimize the expression of pathway genes by implementing plant-derived ATFPs under the inducible control of the GAL1 regulatory system. Production of  $\beta$ -carotene,  $\beta$ -ionone, and NG in *S. cerevisiae* using COMPASS highlights the importance of pathway optimization for the

## DISCUSSION

---

development of microbes capable of industrially reasonable metabolite production. The COMPASS method described here will be of immediate relevance to researchers in synbio and metabolic engineering fields. CaPRedit was also introduced to facilitate integration of inducible, strong plant-derived ATFP using CRISPR-mediated genome editing for enhancing any desired metabolite production. High-level production of  $\beta$ -carotene in the CaPRedit\_FPP 1.0 strain optimized to produce FPP (a  $\beta$ -carotene precursor) by overexpression of only three genes emphasizes the practical application of the CaPRedit approach for other metabolic engineering projects. Future applications of the tools and technologies developed in this study for improving the microbial production of desired products could revolutionize the transition from the laboratory to the marketplace.

## 8 REFERENCES

1. S, L., La, Biologie synthétique. . *A Poinat, Paris* **1912**.
2. Vg, D., Synthetic biology: computational modeling bridging the gap between in vitro and in vivo reactions. *Current Synthetic and Systems Biology* **2015**, *03* (03).
3. Brophy, J. A.; Voigt, C. A., Principles of genetic circuit design. *Nat Methods* **2014**, *11* (5), 508-20.
4. Morgan, S. A.; Nadler, D. C.; Yokoo, R.; Savage, D. F., Biofuel metabolic engineering with biosensors. *Curr Opin Chem Biol* **2016**, *35*, 150-158.
5. Temme, K.; Zhao, D.; Voigt, C. A., Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc Natl Acad Sci U S A* **2012**, *109* (18), 7085-90.
6. Ro, D. K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; Chang, M. C.; Withers, S. T.; Shiba, Y.; Sarpong, R.; Keasling, J. D., Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006**, *440* (7086), 940-3.
7. Karpinski, J.; Hauber, I.; Chemnitz, J.; Schafer, C.; Paszkowski-Rogacz, M.; Chakraborty, D.; Beschorner, N.; Hofmann-Sieber, H.; Lange, U. C.; Grundhoff, A.; Hackmann, K.; Schrock, E.; Abi-Ghanem, J.; Pisabarro, M. T.; Surendranath, V.; Schambach, A.; Lindner, C.; van Lunzen, J.; Hauber, J.; Buchholz, F., Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nat Biotechnol* **2016**, *34* (4), 401-9.
8. Paddon, C. J.; Westfall, P. J.; Pitera, D. J.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M. D.; Tai, A.; Main, A.; Eng, D.; Polichuk, D. R.; Teoh, K. H.; Reed, D. W.; Treynor, T.; Lenihan, J.; Fleck, M.; Bajad, S.; Dang, G.; Dengrove, D.; Diola, D.; Dorin, G.; Ellens, K. W.; Fickes, S.; Galazzo, J.; Gaucher, S. P.; Geistlinger, T.; Henry, R.; Hepp, M.; Horning, T.; Iqbal, T.; Jiang, H.; Kizer, L.; Lieu, B.; Melis, D.; Moss, N.; Regentin, R.; Secrest, S.; Tsuruta, H.; Vazquez, R.; Westblade, L. F.; Xu, L.; Yu, M.; Zhang, Y.; Zhao, L.; Lievens, J.; Covello, P. S.; Keasling, J. D.; Reiling, K. K.; Renninger, N. S.; Newman, J. D., High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **2013**, *496* (7446), 528-32.
9. Shipman, S. L.; Nivala, J.; Macklis, J. D.; Church, G. M., CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* **2017**, *547* (7663), 345-349.
10. H.-J., L., Playing God and the intrinsic value of life: moral problems for synthetic biology? *Sci. Eng. Ethics*. **2012**, *19*, 435-448.
11. Hewett, J. P.; Wolfe, A. K.; Bergmann, R. A.; Stelling, S. C.; Davis, K. L., Human health and environmental risks posed by synthetic biology R&D for energy applications. *Applied Biosafety* **2016**, *21* (4), 177-184.
12. Mattanovich, D.; Sauer, M.; Gasser, B., Yeast biotechnology: teaching the old dog new tricks. *Microb Cell Fact* **2014**, *13* (1), 34.
13. Nielsen, J.; Fussenegger, M.; Keasling, J.; Lee, S. Y.; Liao, J. C.; Prather, K.; Palsson, B., Engineering synergy in biotechnology. *Nat Chem Biol* **2014**, *10* (5), 319-22.
14. Nielsen, J., Production of biopharmaceutical proteins by yeast: advances through metabolic engineering. *Bioengineered* **2013**, *4* (4), 207-11.
15. Ro, D.-K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J., Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006**, *440* (7086), 940-943.
16. Guo, Y.; Dong, J.; Zhou, T.; Auxillos, J.; Li, T.; Zhang, W.; Wang, L.; Shen, Y.; Luo, Y.; Zheng, Y.; Lin, J.; Chen, G. Q.; Wu, Q.; Cai, Y.; Dai, J., YeastFab: the design and construction of standard biological parts for metabolic engineering in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **2015**, *43* (13), e88.
17. Naseri, G.; Balazadeh, S.; Machens, F.; Kamranfar, I.; Messerschmidt, K.; Mueller-Roeber, B., Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth Biol* **2017**, *6*(9):1742-1756.
18. Tsai, C. S.; Kwak, S.; Turner, T. L.; Jin, Y. S., Yeast synthetic biology toolbox and applications for biofuel production. *FEMS Yeast Res* **2015**, *15* (1), 1-15.
19. Purnick, P. E.; Weiss, R., The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* **2009**, *10* (6), 410-22.

## REFERENCES

---

20. Jensen, M. K.; Keasling, J. D., Recent applications of synthetic biology tools for yeast metabolic engineering. *FEMS Yeast Res* **2015**, *15* (1), 1-10.
21. Singh, V., Recent advancements in synthetic biology: current status and challenges. *Gene* **2014**, *535* (1), 1-11.
22. Mishra, D.; Rivera, P. M.; Lin, A.; Del Vecchio, D.; Weiss, R., A load driver device for engineering modularity in biological networks. *Nat Biotechnol* **2014**, *32* (12), 1268-75.
23. Phelan, R. M.; Sachs, D.; Petkiewicz, S. J.; Barajas, J. F.; Blake-Hedges, J. M.; Thompson, M. G.; Reider Apel, A.; Rasor, B. J.; Katz, L.; Keasling, J. D., Development of next generation synthetic biology tools for use in *Streptomyces venezuelae*. *ACS Synth Biol* **2017**, *6* (1), 159-166.
24. Brown, S.; Clastre, M.; Courdavault, V.; O'Connor, S. E., De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc Natl Acad Sci U S A* **2015**, *112* (11), 3205-10.
25. Hochrein, L.; Machens, F.; Gremmels, J.; Schulz, K.; Messerschmidt, K.; Mueller-Roeber, B., AssemblX: a user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Res* **2017**.
26. Anderson, J. C.; Dueber, J. E.; Leguia, M.; Wu, G. C.; Goler, J. A.; Arkin, A. P.; Keasling, J. D., BglBricks: A flexible standard for biological part assembly. *J Biol Eng* **2010**, *4* (1), 1.
27. Mutalik, V. K.; Guimaraes, J. C.; Cambray, G.; Lam, C.; Christoffersen, M. J.; Mai, Q. A.; Tran, A. B.; Paull, M.; Keasling, J. D.; Arkin, A. P.; Endy, D., Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* **2013**, *10* (4), 354-60.
28. Kim, I. K.; Roldao, A.; Siewers, V.; Nielsen, J., A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Res* **2012**, *12* (2), 228-48.
29. Stricker, J.; Cookson, S.; Bennett, M. R.; Mather, W. H.; Tsimring, L. S.; Hasty, J., A fast, robust and tunable synthetic gene oscillator. *Nature* **2008**, *456* (7221), 516-9.
30. Karim, A. S.; Curran, K. A.; Alper, H. S., Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Res* **2013**, *13* (1), 107-16.
31. Bailey, J. E.; Sburlati, A.; Hatzimanikatis, V.; Lee, K.; Renner, W. A.; Tsai, P. S., Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnol Bioeng* **2002**, *79* (5), 568-79.
32. Jones, J. A.; Vernacchio, V. R.; Lachance, D. M.; Lebovich, M.; Fu, L.; Shirke, A. N.; Schultz, V. L.; Cress, B.; Linhardt, R. J.; Koffas, M. A., ePathOptimize: A combinatorial approach for transcriptional balancing of metabolic pathways. *Sci Rep* **2015**, *5*, 11301.
33. Mitsuda, N.; M. Ohme-Takagi, Functional analysis of transcription factors in Arabidopsis. *Plant Cell Physiol* **2009**, *50* (7), 1232-48.
34. T. Meshi, M. I., Plant transcription factors. *Plant Cell Physiol*. **1995**, *36*, 1405-1420.
35. Sakuma, Y.; Liu, Q.; Dubouzet, J. G.; Abe, H.; Shinozaki, K.; Yamaguchi-Shinozaki, K., DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem Biophys Res Commun* **2002**, *290* (3), 998-1009.
36. Schindler, U., Terzaghi, W., Beckmann, H., Kadesch, T. and Cashmore, A.R., DNA binding site preferences and transcriptional activation properties of the Arabidopsis transcription factor GBF1. *EMBO J*. **1992**, *11*, 1275 – 1289.
37. Döring, P., Treuter, E., Kistner, C., Lyck, R., Chen, A. and Nover, L., The role of AHA motifs in the activator function of tomato heat stress transcription factors HsfA1 and HsfA2. *Plant Cell* **2000**, *12*, 265 – 278.
38. Gaj, T.; Gersbach, C. A.; Barbas, C. F., 3rd, ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **2013**, *31* (7), 397-405.
39. Redden, H.; Alper, H. S., The development and characterization of synthetic minimal yeast promoters. *Nat Commun* **2015**, *6*, 7810.
40. Rajkumar, A. S.; Denervaud, N.; Maerkl, S. J., Mapping the fine structure of a eukaryotic promoter input-output function. *Nat Genet* **2013**, *45* (10), 1207-15.
41. Partow, S.; Siewers, V.; Bjorn, S.; Nielsen, J.; Maury, J., Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast* **2010**, *27* (11), 955-64.
42. Kabadi, A. M.; Gersbach, C. A., Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression. *Methods* **2014**, *69* (2), 188-97.
43. Purcell, O.; Peccoud, J.; Lu, T. K., Rule-based design of synthetic transcription factors in eukaryotes. *ACS Synth Biol* **2014**, *3* (10), 737-44.



## REFERENCES

---

44. Jensen, E. D.; Ferreira, R.; Jakociunas, T.; Arsovska, D.; Zhang, J.; Ding, L.; Smith, J. D.; David, F.; Nielsen, J.; Jensen, M. K.; Keasling, J. D., Transcriptional reprogramming in yeast using dCas9 and combinatorial gRNA strategies. *Microb Cell Fact* **2017**, *16* (1), 46.
45. Khalil, A. S.; Lu, T. K.; Bashor, C. J.; Ramirez, C. L.; Pyenson, N. C.; Joung, J. K.; Collins, J. J., A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **2012**, *150* (3), 647-58.
46. Garg, A.; Lohmueller, J. J.; Silver, P. A.; Armel, T. Z., Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res* **2012**, *40* (15), 7584-95.
47. Wu, G.; Yan, Q.; Jones, J. A.; Tang, Y. J.; Fong, S. S.; Koffas, M. A. G., Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol* **2016**, *34* (8), 652-664.
48. Jones, K. L.; Kim, S. W.; Keasling, J. D., Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab Eng* **2000**, *2* (4), 328-38.
49. Tyo, K. E.; Ajikumar, P. K.; Stephanopoulos, G., Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat Biotechnol* **2009**, *27* (8), 760-5.
50. Jakociunas, T.; Rajkumar, A. S.; Zhang, J.; Arsovska, D.; Rodriguez, A.; Jendresen, C. B.; Skjodt, M. L.; Nielsen, A. T.; Borodina, I.; Jensen, M. K.; Keasling, J. D., CasEMBLR: Cas9-facilitated multiloci genomic integration of in vivo assembled DNA parts in *Saccharomyces cerevisiae*. *ACS Synth Biol* **2015**, *4* (11), 1226-34.
51. Cuperus, J. T.; Lo, R. S.; Shumaker, L.; Proctor, J.; Fields, S., A tetO Toolkit To Alter Expression of Genes in *Saccharomyces cerevisiae*. *ACS Synth Biol* **2015**, *4* (7), 842-52.
52. Barrangou, R., The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Curr Opin Immunol* **2015**, *32*, 36-41.
53. Koonin, E. V.; Makarova, K. S., CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol* **2013**, *10* (5), 679-86.
54. Mercy, G.; Mozziconacci, J.; Scolari, V. F.; Yang, K.; Zhao, G.; Thierry, A.; Luo, Y.; Mitchell, L. A.; Shen, M.; Shen, Y.; Walker, R.; Zhang, W.; Wu, Y.; Xie, Z.-x.; Luo, Z.; Cai, Y.; Dai, J.; Yang, H.; Yuan, Y.-J.; Boeke, J. D.; Bader, J. S.; Muller, H.; Koszul, R., 3D organization of synthetic and scrambled chromosomes. *Science* **2017**, *355* (6329).
55. Anthony, J. R.; Anthony, L. C.; Nowroozi, F.; Kwon, G.; Newman, J. D.; Keasling, J. D., Optimization of the mevalonate-based isoprenoid biosynthetic pathway in *Escherichia coli* for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metab Eng* **2009**, *11* (1), 13-9.
56. Bao, Z.; Xiao, H.; Liang, J.; Zhang, L.; Xiong, X.; Sun, N.; Si, T.; Zhao, H., Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS Synth Biol* **2015**, *4* (5), 585-94.
57. J Lian, M. H., H. Zhao, Advancing metabolic engineering of *Saccharomyces cerevisiae* using the CRISPR/Cas system. *Biotechnol J* **2018**.
58. Zalatan, J. G.; Lee, M. E.; Almeida, R.; Gilbert, L. A.; Whitehead, E. H.; La Russa, M.; Tsai, J. C.; Weissman, J. S.; Dueber, J. E.; Qi, L. S.; Lim, W. A., Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **2015**, *160* (1-2), 339-50.
59. Gao, X.; Tsang, J. C.; Gaba, F.; Wu, D.; Lu, L.; Liu, P., Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Res* **2014**, *42* (20), e155.
60. Buijs, N. A.; Siewers, V.; Nielsen, J., Advanced biofuel production by the yeast *Saccharomyces cerevisiae*. *Current Opinion in Chemical Biology* **2013**, *17* (3), 480-8.
61. Nielsen, J.; Keasling, J. D., Engineering Cellular Metabolism. *Cell* **2016**, *164* (6), 1185-97.
62. Vanegas, K.G.; Lehka, B.J.; Mortensen U.H., SWITCH: a dynamic CRISPR tool for genome engineering and metabolic pathway control for cell factory construction in *Saccharomyces cerevisiae*. *Microb Cell Fact.* **2017**, 16:25.
63. Deaner, M.; Alper, H.S., Systematic testing of enzyme perturbation sensitivities via graded dCas9 modulation in *Saccharomyces cerevisiae*. *Metab. Eng.* **2017**, *40*, 14-22.
64. Deaner, M.; Mejia, J.; Alper, H.S., Enabling graded and large-scale multiplex of desired genes using a dual-mode dCas9 activator in *Saccharomyces cerevisiae*. *ACS Synth Biol* **2017**, *6*, 1931-1943.
65. Lian, J.; HamediRad, M.; Hu, S.; Zhao, H., Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. *Nat Commun* **2017**, *8* (1), 1688.

## REFERENCES

---

66. Jakociunas, T.; Bonde, I.; Herrgard, M.; Harrison, S. J.; Kristensen, M.; Pedersen, L. E.; Jensen, M. K.; Keasling, J. D., Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. *Metab Eng* **2015**, *28*, 213-22.
67. Keasling, J. D., Gene-expression tools for the metabolic engineering of bacteria. *Trends Biotechnol* **1999**, *17* (11), 452-60.
68. Cohen, S. N., Chang, A. C., Boyer, H. W., Helling, R. B., Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci.* **1973**, *70*, 3240–3244.
69. Woolston, B. M.; Edgar, S.; Stephanopoulos, G., Metabolic engineering: past and future. *Annu Rev Chem Biomol Eng* **2013**, *4*, 259-88.
70. Frazzetto, G., White biotechnology. *EMBO Rep.* **2003**, *4*, 835–837.
71. Runguphan, W.; Keasling, J. D., Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab Eng* **2014**, *21*, 103-13.
72. Newman, J. D.; Marshall, J.; Chang, M.; Nowroozi, F.; Paradise, E.; Pitera, D.; Newman, K. L.; Keasling, J. D., High-level production of amorpho-4,11-diene in a two-phase partitioning bioreactor of metabolically engineered *Escherichia coli*. *Biotechnol Bioeng* **2006**, *95* (4), 684-91.
73. Siddiqui, M. S.; Thodey, K.; Trenchard, I.; Smolke, C. D., Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res* **2012**, *12* (2), 144-70.
74. Hong, K. K.; Nielsen, J., Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cell Mol Life Sci* **2012**, *69* (16), 2671-90.
75. Cazzonelli, C. I.; Cuttriss, A. J.; Cossetto, S. B.; Pye, W.; Crisp, P.; Whelan, J.; Finnegan, E. J.; Turnbull, C.; Pogson, B. J., Regulation of carotenoid composition and shoot branching in *Arabidopsis* by a chromatin modifying histone methyltransferase, SDG8. *Plant Cell* **2009**, *21* (1), 39-53.
76. Walter, M. H.; Strack, D., Carotenoids and their cleavage products: biosynthesis and functions. *Nat Prod Rep* **2011**, *28* (4), 663-92.
77. Mortensen, A., Carotenoids and other pigments as natural colorants. *Pure and Applied Chemistry* **2006**, *78* (8).
78. Kang, A.; Meadows, C. W.; Canu, N.; Keasling, J. D.; Lee, T. S., High-throughput enzyme screening platform for the IPP-bypass mevalonate pathway for isopentenol production. *Metab Eng* **2017**, *41*, 125-134.
79. M. E. Basson, M. T., J. Finer-Moore, R. M. Stroud, J. Rine, Structural and functional conservation between yeast and human 3-hydroxy-3-methylglutaryl coenzyme A reductases, the rate-limiting enzyme of sterol biosynthesis. *Mol Cell Biol* **1998**, *8*, 3797–3808.
80. Martin, V. J.; Yoshikuni, Y.; Keasling, J. D., The in vivo synthesis of plant sesquiterpenes by *Escherichia coli*. *Biotechnol Bioeng* **2001**, *75* (5), 497-503.
81. Scalcinati, G.; Partow, S.; Siewers, V.; Schalk, M.; Daviet, L.; Nielsen, J., Combined metabolic engineering of precursor and co-factor supply to increase alpha-santalene production by *Saccharomyces cerevisiae*. *Microb Cell Fact* **2012**, *11*, 117.
82. Lopez, J.; Essus, K.; Kim, I. K.; Pereira, R.; Herzog, J.; Siewers, V.; Nielsen, J.; Agosin, E., Production of beta-ionone by combined expression of carotenogenic and plant CCD1 genes in *Saccharomyces cerevisiae*. *Microb Cell Fact* **2015**, *14*, 84.
83. Falcone Ferreyra, M. L.; Rius, S. P.; Casati, P., Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci* **2012**, *3*, 222.
84. K.G. Ryan, E. E. S., C. Winefield, K.R. Markham, Flavonoids and UV photoprotection in *Arabidopsis* mutants. *Z Naturforsch [C]* **2001**, *56*, 745-754.
85. Peters, J. R.; Keasling, R.; Brown, S. D.; Pond, B. B., Quantification of synthetic cathinones in rat brain using HILIC-ESI-MS/MS. *J Anal Toxicol* **2016**, *40* (9), 718-725.
86. Fowler, Z. L.; Koffas, M. A., Biosynthesis and biotechnological production of flavanones: current state and perspectives. *Appl Microbiol Biotechnol* **2009**, *83* (5), 799-808.
87. Koopman F.; Beekwilder J.; Crimi B.; Houwelingen A.; Hall R. D.; Bosch D.; Maris J.; Pronk J. T.; Daran, J.-M., De novo production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*. *Microbial Cell Factories* **2012**, *11*, 155.
88. E. Leonard, Y. Y., Z. L. Fowler, Z. Li, C.G. Lim, K.H. Lim, M.A.G. Koffas, Strain Improvement of Recombinant *Escherichia coli* for Efficient Production of Plant Flavonoids. *Mol. Pharmaceutics* **2008**, *5*, 257–265.
89. Zhang, W.; Liu, H.; Li, X.; Liu, D.; Dong, X.-T.; Li, F.-F.; Wang, E.-X.; Li, B.-Z.; Yuan, Y.-J., Production of naringenin from D-xylose with co-culture of *E. coli* and *S. cerevisiae*. *Engineering in Life Sciences* **2017**, *17* (9), 1021-1029.

## REFERENCES

90. J P Pinto, R. P., J Cardoso, I Rocha, M Rocha, TNA4OptFlux – a software tool for the analysis of strain optimization strategies. *BMC Research Notes* **2013**, *6*, 175-188.
91. Ding, M. Z.; Yan, H. F.; Li, L. F.; Zhai, F.; Shang, L. Q.; Yin, Z.; Yuan, Y. J., Biosynthesis of Taxadiene in *Saccharomyces cerevisiae* : selection of geranylgeranyl diphosphate synthase directed by a computer-aided docking strategy. *PLoS One* **2014**, *9* (10), e109348.
92. Hester, J. B.; Rudzik, A. D.; Keasling, H. H.; Veldkamp, W., 4'-Fluoro-4-(1,4,5,6-tetrahydroazepinol[4,5-b]indol-3(2H)-yl)butyrophenones. *J Med Chem* **1970**, *13* (1), 23-6.
93. Latimer, L. N.; Dueber, J. E., Iterative optimization of xylose catabolism in *Saccharomyces cerevisiae* using combinatorial expression tuning. *Biotechnol Bioeng* **2017**, *114* (6), 1301-1309.
94. Du, J.; Yuan, Y.; Si, T.; Lian, J.; Zhao, H., Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res* **2012**, *40* (18), e142.
95. Rodrigues, R., Improving microbial chemical production strains through transcriptional regulatory network rewiring. *PhD thesis, Imperial College London, Department of Life Sciences* **2015**.
96. Subtil, T.; Boles, E., Competition between pentoses and glucose during uptake and catabolism in recombinant *Saccharomyces cerevisiae*. *Biotechnol Biofuels* **2012**, *5*, 14.
97. Li, Y. J.; Wang, M. M.; Chen, Y. W.; Wang, M.; Fan, L. H.; Tan, T. W., Engineered yeast with a CO<sub>2</sub>-fixation pathway to improve the bio-ethanol production from xylose-mixed sugars. *Sci Rep* **2017**, *7*, 43875.
98. J Barrios-Gonzalez , F. J. F., A. Tomasini, Microbial secondary metabolites production and strain improvement. *Indian Journal of Biotechnology* **2003**, *2* , 322-333.
99. Thodey, K.; Galanie, S.; Smolke, C. D., A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat Chem Biol* **2014**, *10* (10), 837-44.
100. Pflieger, B. F.; Pitera, D. J.; Smolke, C. D.; Keasling, J. D., Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* **2006**, *24* (8), 1027-32.
101. Bond-Watts, B. B.; Bellerose, R. J.; Chang, M. C., Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nat Chem Biol* **2011**, *7* (4), 222-7.
102. Shen, C. R.; Lan, E. I.; Dekishima, Y.; Baez, A.; Cho, K. M.; Liao, J. C., Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Appl Environ Microbiol* **2011**, *77* (9), 2905-15.
103. Bujara, M.; Panke, S., Engineering in complex systems. *Curr Opin Biotechnol* **2010**, *21*, 586–591.
104. Bailey, A. S.; Stanescu, S.; Yeasting, R. A.; Ebraheim, N. A.; Jackson, W. T., Anatomic relationships of the cervicothoracic junction. *Spine (Phila Pa 1976)* **1995**, *20* (13), 1431-9.
105. Jin, P.; Kang, Z.; Zhang, J.; Zhang, L.; Du, G.; Chen, J., Combinatorial evolution of enzymes and synthetic pathways using one-step PCR. *ACS Synth Biol* **2016**, *5* (3), 259-68.
106. Gill, R. T., Enabling inverse metabolic engineering through genomics. *Current Opinion in Biotechnology* **2003**, *14* (5), 484-490.
107. Gill, R. T.; Wildt, S.; Yang, Y. T.; Ziesman, S.; Stephanopoulos, G., Genome-wide screening for trait conferring genes using DNA microarrays. *Proc Natl Acad Sci U S A* **2002**, *99* (10), 7033-8.
108. Zhang, F.; Keasling, J., Biosensors and their applications in microbial metabolic engineering. *Trends Microbiol* **2011**, *19* (7), 323-9.
109. Skretas, G.; Kolisis, F. N., Combinatorial approaches for inverse metabolic engineering applications. *Comput Struct Biotechnol J* **2012**, *3*, e201210021.
110. Jeschek, M.; Gerngross, D.; Panke, S., Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nat Commun* **2016**, *7*, 11163.
111. D'Ambrosio, V.; Jensen, M. K., Lighting up yeast cell factories by transcription factor-based biosensors. *FEMS Yeast Res* **2017**, *17* (7).
112. Siedler, S.; Stahlhut, S. G.; Malla, S.; Maury, J.; Neves, A. R., Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into *Escherichia coli*. *Metab Eng* **2014**, *21*, 2-8.
113. 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.; Siedler, S.; Borodina, I.; Jensen, M. K.; Keasling, J. D., Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nat Chem Biol* **2016**, *12* (11), 951-958.
114. Lai, E. C., RNA sensors and riboswitches: self-regulating messages. *Current Biology* **2003**, *13* (7), R285-R291.
115. Harbaugh, S. V.; Goodson, M. S.; Dillon, K.; Zabarnick, S.; Kelley-Loughnane, N., Riboswitch-based reversible dual color sensor. *ACS Synth Biol* **2017**, *6* (5), 766-781.

## REFERENCES

---

116. Lin, J. L.; Wagner, J. M.; Alper, H. S., Enabling tools for high-throughput detection of metabolites: Metabolic engineering and directed evolution applications. *Biotechnol Adv* **2017**, *35* (8), 950-970.
117. Friedland, A. E.; Lu, T. K.; Wang, X.; Shi, D.; Church, G.; Collins, J. J., Synthetic gene networks that count. *Science* **2009**, *324* (5931), 1199-202.
118. Beekwilder, J.; van Leeuwen, W.; van Dam, N. M.; Bertossi, M.; Grandi, V.; Mizzi, L.; Soloviev, M.; Szabados, L.; Molthoff, J. W.; Schipper, B.; Verbocht, H.; de Vos, R. C.; Morandini, P.; Aarts, M. G.; Bovy, A., The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. *PLoS One* **2008**, *3* (4), e2068.
119. Tsuruta, H.; Paddon, C. J.; Eng, D.; Lenihan, J. R.; Horning, T.; Anthony, L. C.; Regentin, R.; Keasling, J. D.; Renninger, N. S.; Newman, J. D., High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. *PLoS One* **2009**, *4* (2), e4489.
120. Szymczak, A. L.; Workman, C. J.; Wang, Y.; Vignali, K. M.; Dilioglou, S.; Vanin, E. F.; Vignali, D. A., Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat Biotechnol* **2004**, *22* (5), 589-94.
121. Mizuguchi, H.; Xu, Z.; Ishii-Watabe, A.; Uchida, E.; Hayakawa, T., IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. *Mol Ther* **2000**, *1* (4), 376-82.
122. Beekwilder, J.; van Rossum, H. M.; Koopman, F.; Sonntag, F.; Buchhaupt, M.; Schrader, J.; Hall, R. D.; Bosch, D.; Pronk, J. T.; van Maris, A. J.; Daran, J. M., Polycistronic expression of a beta-carotene biosynthetic pathway in *Saccharomyces cerevisiae* coupled to beta-ionone production. *J Biotechnol* **2014**, *192 Pt B*, 383-92.
123. Verwaal, R.; Wang, J.; Meijnen, J. P.; Visser, H.; Sandmann, G.; van den Berg, J. A.; van Ooyen, A. J., High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive transformation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. *Appl Environ Microbiol* **2007**, *73* (13), 4342-50.
124. Perez-Rodriguez, P.; Riano-Pachon, D. M.; Correa, L. G.; Rensing, S. A.; Kersten, B.; Mueller-Roeber, B., PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res* **2010**, *38* (Database issue), D822-7.
125. Yamasaki, K.; Kigawa, T.; Seki, M.; Shinozaki, K.; Yokoyama, S., DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends Plant Sci* **2013**, *18* (5), 267-76.
126. Allu, A. D.; Brotman, Y.; Xue, G. P.; Balazadeh, S., Transcription factor ANAC032 modulates JA/SA signalling in response to *Pseudomonas syringae* infection. *EMBO Rep* **2016**, *17* (11), 1578-1589.
127. Sayou, C.; Monniaux, M.; Nanao, M. H.; Moyroud, E.; Brockington, S. F.; Thevenon, E.; Chahtane, H.; Warthmann, N.; Melkonian, M.; Zhang, Y.; Wong, G. K.; Weigel, D.; Parcy, F.; Dumas, R., A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* **2014**, *343* (6171), 645-8.
128. Yanagisawa, S., The Dof family of plant transcription factors. *TRENDS in Plant Science* **2002**, *7* 555-560.
129. Prouse, M. B.; Campbell, M. M., Interactions between the R2R3-MYB transcription factor, AtMYB61, and target DNA binding sites. *PLoS One* **2013**, *8* (5), e65132.
130. Ciolkowski, I.; Wanke, D.; Birkenbihl, R. P.; Somssich, I. E., Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. *Plant Mol Biol* **2008**, *68* (1-2), 81-92.
131. Nole-Wilson, S.; Krizek, B. A., DNA binding properties of the Arabidopsis floral development protein Aintegumenta. *Nucleic Acids Research* **2000**, *28* (21), 4076-4082.
132. Kagaya, Y.; Ohmiya, K.; Hattori, T., RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Research* **1999**, *Vol. 27* (No. 2), 470-478.
133. Kim, J. S.; Mizoi, J.; Kidokoro, S.; Maruyama, K.; Nakajima, J.; Nakashima, K.; Mitsuda, N.; Takiguchi, Y.; Ohme-Takagi, M.; Kondou, Y.; Yoshizumi, T.; Matsui, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K., Arabidopsis growth-regulating factor7 functions as a transcriptional repressor of abscisic acid- and osmotic stress-responsive genes, including DREB2A. *Plant Cell* **2012**, *24* (8), 3393-405.
134. Kim, J. H.; Choi, D.; Kende, H., The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. *Plant J* **2003**, *36*, 94-104.
135. Wu, A.; Allu, A. D.; Garapati, P.; Siddiqui, H.; Dortay, H.; Zanol, M. I.; Asensi-Fabado, M. A.; Munne-Bosch, S.; Antonio, C.; Tohge, T.; Fernie, A. R.; Kaufmann, K.; Xue, G. P.; Mueller-Roeber, B.;

## REFERENCES

---

- Balazadeh, S., JUNGBRUNNEN1, a reactive oxygen species-responsive NAC transcription factor, regulates longevity in Arabidopsis. *Plant Cell* **2012**, *24* (2), 482-506.
136. Matallana-Ramirez, L. P.; Rauf, M.; Farage-Barhom, S.; Dortay, H.; Xue, G. P.; Droge-Laser, W.; Lers, A.; Balazadeh, S.; Mueller-Roeber, B., NAC transcription factor ORE1 and senescence-induced bifunctional nuclease1 (BFN1) constitute a regulatory cascade in Arabidopsis. *Mol Plant* **2013**, *6* (5), 1438-52.
137. Garapati, P.; Xue, G. P.; Munne-Bosch, S.; Balazadeh, S., Transcription factor ATAF1 in Arabidopsis promotes senescence by direct regulation of key chloroplast maintenance and senescence transcriptional cascades. *Plant Physiol* **2015**, *168* (3), 1122-39.
138. Christianson, J. A.; Wilson, I. W.; Llewellyn, D. J.; Dennis, E. S., The low-oxygen-induced NAC domain transcription factor ANAC102 affects viability of Arabidopsis seeds following low-oxygen treatment. *Plant Physiol* **2009**, *149* (4), 1724-38.
139. Krizek, B. A.; Sulli, C., Mapping sequences required for nuclear localization and the transcriptional activation function of the Arabidopsis protein Aintegumenta. *Planta* **2006**, *224* (3), 612-21.
140. Inze, A.; Vanderauwera, S.; Hoerberichts, F. A.; Vandorpe, M.; Van Gaever, T.; Van Breusegem, F., A subcellular localization compendium of hydrogen peroxide-induced proteins. *Plant Cell Environ* **2012**, *35* (2), 308-20.
141. Stender, E. G.; O'Shea, C.; Skriver, K., Subgroup-specific intrinsic disorder profiles of Arabidopsis NAC transcription factors: Identification of functional hotspots. *Plant Signal Behav* **2015**, *10* (6), e1010967.
142. Kelwick, R.; MacDonald, J. T.; Webb, A. J.; Freemont, P., Developments in the tools and methodologies of synthetic biology. *Front Bioeng Biotechnol* **2014**, *2*, 60.
143. Julleson, D.; David, F.; Pflieger, B.; Nielsen, J., Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnol Adv* **2015**, *33* (7), 1395-402.
144. Chubukov, V.; Mukhopadhyay, A.; Petzold, C. J.; Keasling, J. D.; Martin, H. G., Synthetic and systems biology for microbial production of commodity chemicals. *NPJ Syst Biol Appl* **2016**, *2*, 16009.
145. Nakamura, C. E.; Whited, G. M., Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion in Biotechnology* **2003**, *14* (5), 454-459.
146. Yim, H.; Haselbeck, R.; Niu, W.; Pujol-Baxley, C.; Burgard, A.; Boldt, J.; Khandurina, J.; Trawick, J. D.; Osterhout, R. E.; Stephen, R.; Estadilla, J.; Teisan, S.; Schreyer, H. B.; Andrae, S.; Yang, T. H.; Lee, S. Y.; Burk, M. J.; Van Dien, S., Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat Chem Biol* **2011**, *7* (7), 445-52.
147. Scalcinati, G.; Knuf, C.; Partow, S.; Chen, Y.; Maury, J.; Schalk, M.; Daviet, L.; Nielsen, J.; Siewers, V., Dynamic control of gene expression in *Saccharomyces cerevisiae* engineered for the production of plant sesquiterpene alpha-santalene in a fed-batch mode. *Metab Eng* **2012**, *14* (2), 91-103.
148. Kerkhoven, E. J.; Lahtvee, P. J.; Nielsen, J., Applications of computational modeling in metabolic engineering of yeast. *FEMS Yeast Res* **2015**, *15* (1), 1-13.
149. Jin, Y. S.; Stephanopoulos, G., Multi-dimensional gene target search for improving lycopene biosynthesis in *Escherichia coli*. *Metab Eng* **2007**, *9* (4), 337-47.
150. Lechner, A.; Brunk, E.; Keasling, J. D., The Need for Integrated Approaches in Metabolic Engineering. *Cold Spring Harb Perspect Biol* **2016**, *8* (11).
151. Lv, X.; Wang, F.; Zhou, P.; Ye, L.; Xie, W.; Xu, H.; Yu, H., Dual regulation of cytoplasmic and mitochondrial acetyl-CoA utilization for improved isoprene production in *Saccharomyces cerevisiae*. *Nat Commun* **2016**, *7*, 12851.
152. Mitchell, L. A.; Chuang, J.; Agmon, N.; Khunsriraksakul, C.; Phillips, N. A.; Cai, Y.; Truong, D. M.; Veerakumar, A.; Wang, Y.; Mayorga, M.; Blomquist, P.; Sadda, P.; Trueheart, J.; Boeke, J. D., Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Res* **2015**, *43* (13), 6620-30.
153. Wang, H. H.; Isaacs, F. J.; Carr, P. A.; Sun, Z. Z.; Xu, G.; Forest, C. R.; Church, G. M., Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **2009**, *460* (7257), 894-898.
154. DiCarlo, J. E.; Conley, A. J.; Penttila, M.; Jantti, J.; Wang, H. H.; Church, G. M., Yeast oligo-mediated genome engineering (YOGE). *ACS Synth Biol* **2013**, *2* (12), 741-9.
155. Baek, C. H.; Chesnut, J.; Katzen, F., Positive selection improves the efficiency of DNA assembly. *Anal Biochem* **2015**, *476*, 1-4.
156. Zhang, Y.; Werling, U.; Edelman, W., SLICE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res* **2012**, *40* (8), e55.

## REFERENCES

---

157. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **2009**, *6* (5), 343-345.
158. Zhang, Y.; Werling, U.; Edlmann, W., SLICE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Research* **2012**, *40* (8), e55.
159. Kouprina, N.; Larionov, V., TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nature Reviews Genetics* **2006**, *7* (10), 805-812.
160. Kouprina, N.; Larionov, V., Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* **2016**, *125* (4), 621-32.
161. Sternberg, N.; Hamilton, D., Bacteriophage P1 site-specific recombination: I. Recombination between loxP sites. *J. Mol. Biol* **1981**, *150*, 467-486.
162. Meyer, A. J.; Ellefson, J. W.; Ellington, A. D., Library generation by gene shuffling. *Curr Protoc Mol Biol* **2014**, *105*, Unit 15 12.
163. Shimizu-Sato, S.; Huq, E.; Tepperman, J. M.; Quail, P. H., A light-switchable gene promoter system. *Nat Biotechnol* **2002**, *20* (10), 1041-4.
164. Christie, J. M.; Gawthorne, J.; Young, G.; Fraser, N. J.; Roe, A. J., LOV to BLUF: flavoprotein contributions to the optogenetic toolkit. *Mol Plant* **2012**, *5* (3), 533-44.
165. Ghosh, A.; Ando, D.; Gin, J.; Runguphan, W.; Denby, C.; Wang, G.; Baidoo, E. E.; Shymansky, C.; Keasling, J. D.; Garcia Martin, H., <sup>13</sup>C Metabolic flux analysis for systematic metabolic engineering of *S. cerevisiae* for Overproduction of Fatty Acids. *Front Bioeng Biotechnol* **2016**, *4*, 76.
166. Breslow, D. K.; Cameron, D. M.; Collins, S. R.; Schuldiner, M.; Stewart-Ornstein, J.; Newman, H. W.; Braun, S.; Madhani, H. D.; Krogan, N. J.; Weissman, J. S., A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat Methods* **2008**, *5* (8), 711-8.

**Personal details**

**Email:** naseri@uni-potsdam.de; gita\_naseri@yahoo.com

**Nationality:** Iran

**Residency status:** Germany

**Education**

**September 2013- now** Ph.D. student in Synthetic Biology, Junior Research Group Cell2Fab, University of Potsdam, Germany

**2006-2009** MSc. Plant Biotechnology, University of Guilan, Iran

**2000-2005** BSc. Plant Pathology, Agricultural Sciences and Natural Resources University of Gorgan, Iran

**1999-2000** Pre-University of Applied Science, Velayat-e-Faghi, Iran

**1996-1999** Diploma of Applied Science, High School of Razi, Iran

**Professional experience**

**2010-2013** Technical manager confirmed by the Ministry of Agriculture, Laboratory of Plant Protection Clinic, Iran

**Publications**

1. ARTICLE: Naseri G., Balazadeh S., Machens F., Kamranfar I., Messerschmidt K., Mueller-Roeber B. (2017). Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 2017, 6, 1742–1756. doi: 10.1021/acssynbio.7b00094.
2. ARTICLE: Naseri G., Behrend J., Rieper L., Mueller-Roeber B. COMPASS: Rapid combinatorial optimization of biochemical pathways based on artificial transcription factors. Submitted to *Nature Methods* (Status: Under consideration).
3. ARTICLE: Naseri G., Behrend J., Schulz K., Messerschmidt K. CaPRedit – CRISPR / Plant-derived Regulator-mediated Genome Editing for Redirection of Flux through the FPP branch-point in yeast. Manuscript in preparation.

**Training courses**

**2016**

1. Workshop “Connaught Summer Institute on Synthetic Biology” 6-10 June. IMPACT CENTER, University of Toronto, Toronto, Canada
2. “Data analysis with MATLAB software” course. 8-12 February. Potsdam, Germany

**2015**

EMBO Practical Course “Synthetic biology in action” 8–20 June. Heidelberg, Germany

**Conferences**

1. Naseri G., Balazadeh S., Machens F., Kamranfar I., Messerschmidt K., Mueller-Roeber B. 25 – 27. September, 2017. Orthogonal regulation of gene expression in yeast using plant-derived transcription factors. 17<sup>th</sup> Euro Biotechnology Congress. Berlin, Germany (oral presentation)
2. Naseri G., Machens F., Mueller-Roeber B., Messerschmidt K. 28. June- 3. July, 2015. Characterization platform for evaluation of heterologous transcription factors and artificial promoters for synthetic biosystems. Gordon Research Conference on Synthetic Biology. Newry, United States of America (poster presentation)
3. Naseri G., Machens F., Mueller-Roeber B., Messerschmidt K. 01.- 04. Juni, 2014. Generating a library of plant-derived transcription factors and promoters for synthetic biosystems. Latin-American Conference on Knowledge and Technology Transfer & Biotechnology and Life Science. Potsdam, Germany (poster presentation)
4. Naseri G., Sohani M. M., Pourmasalehgoo A. and Zamani A.H. 27. - 29. May, 2009. Isolation and cloning of thaumatin-like protein gene for studying the mechanism of resistance to sheath blight (*Rhizoctonia solani*). The 5<sup>th</sup> student’s congress of agricultural sciences and natural resources. Rasht, Iran (poster presentation)
5. Naseri G., Sohani M. M. and Pourmasalehgoo A. 28. – 29. January, 2009. *In planta* transformation of rice by using *Agrobacterium tumefaciens*. The 13th national rice congress “Challenges facing rice self-sufficiently”. Rasht, Iran (poster presentation)
6. Zamani A.H., Sohani M. M., Hatamzadeh A., Afsharifar A., Mirzai M., Pourmasalehgoo A. and Naseri G. 13. – 16. July, 2009. Cloning of the synthetic sequence of Citrus tristeza virus for inducing resistance to this disease in citrus. The 6th Horticulture Congress. Rasht, Iran (poster presentation)

**Honors and awards****2017**

- Young scientist introduced by the ACS Synthetic Biology journal

**2016**

- PoGS scholarship (completion grant) from the University of Potsdam, Germany



**2013**

- CIMO fellowship awarded to do PhD, Finland
- DAAD scholarship awarded to do PhD, Germany

**2009**

- Ranked 1st among all post-graduate students of the University of Guilan-GPA: 1.37

**2000**

- Highest pre-university GPA in the city of Tonekabon, Iran - GPA: 1.12

**1999**

- Highest high school diploma GPA in the city of Tonekabon, Iran - GPA: 1.11

**Certifications**

**2012**

- Certification for agricultural and natural resource engineering society. Iran

**2010**

- Certificate for Technical and vocational skills. Iran
- Certificate for completion of agricultural engineering graduates ' internship projects. Iran

**Language**

- Farsi: Native
- English: Fluent

### SCIENTIFIC ACHIEVEMENTS

#### PUBLICATIONS

1. ARTICLE: Naseri G., Balazadeh S., Machens F., Kamranfar I., Messerschmidt K., Mueller-Roeber B. (2017). Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 2017, 6, 1742–1756. doi: 10.1021/acssynbio.7b00094.
2. ARTICLE: Naseri G., Behrend J., Rieper L., Mueller-Roeber B. COMPASS: Rapid combinatorial optimization of biochemical pathways based on artificial transcription factors. Manuscript submitted to Nature Communicaton (status: under consideration).
3. ARTICLE: Naseri G., Behrend J., Schulz K., Messerschmidt K. CaPRedit – CRISPR / Plant-derived Regulator-mediated Genome Editing for Redirection of Flux through the FPP branch-point in yeast. Manuscript in preparation.

#### CONFERENCE CONTRIBUTIONS

1. Naseri G., Balazadeh S., Machens F., Kamranfar I., Messerschmidt K., Mueller-Roeber B. 25 – 27. September, 2017. Orthogonal regulation of gene expression in yeast using plant-derived transcription factors. 17<sup>th</sup> Euro Biotechnology Congress. Berlin, Germany (oral presentation)
2. Naseri G., Machens F., Mueller-Roeber B., Messerschmidt K. 28. June- 3. July, 2015. Characterization platform for evaluation of heterologous transcription factors and artificial promoters for synthetic biosystems. Gordon Research Conference on Synthetic Biology. Newry, United States of America (poster presentation)
3. Naseri G., Machens F., Mueller-Roeber B., Messerschmidt K. 01.- 04. Juni, 2014. Generating a library of plant-derived transcription factors and promoters for synthetic biosystems. Latin-American Conference on Knowledge and Technology Transfer & Biotechnology and Life Science. Potsdam, Germany (poster presentation)

## DECLARATION

---

### **DECLARATIONS**

#### **DECLARATION OF INDEPENDENT WORK**

I hereby confirm that this Ph.D. thesis is the result of my work. Information derived from the work of others has been acknowledged in the text and the references are given. I declare that this work has not been submitted for any other degree previously, neither at the University of Potsdam nor at any other university.

Potsdam, 15<sup>th</sup> May, 2018

Gita Naseri

### **DECLARATION OF OWN CONTRIBUTIONS TO SINGLE PUBLICATIONS**

#### **Plant-derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces cerevisiae***

Gita Naseri, Salma Balazadeh, Fabian Machens, Iman Kamranfar, Katrin Messerschmidt, Bernd Mueller-Roeber

#### **Author contributions**

GN and BMR designed the experiments, with contributions from KM and FM. SB provided information about the binding sites of transcription factors and advice for the analysis of transcription factor – DNA interactions. IK supported the ANAC102-related experiments. FM developed the pFM series of plasmids. GN performed the experiments and analyzed the data. GN, SB and BMR wrote the manuscript, which was proofread by all authors. All authors take full responsibility for the content of the paper.