Development of a Humanized and Genomically Scramblable Synthetic Neochromosome Platform for Coenzyme A Biosynthesis Study

Jyue-Shi Yang [1,2]

[1]Department of Biological Engineering, Massachusetts Institute of Technology

[2]Manchester Institute of Biotechnology, The University of Manchester

Supervised by Prof. Patrick (Yizhi) Cai from Manchester Institute of Biotechnology

Table of Contents

Abstract

This research developed a humanized, independently scramblable synthetic neochromosome platform in Saccharomyces cerevisiae, utilizing the essential Coenzyme A (CoA) biosynthesis pathway as a model to investigate metabolic plasticity and cross-kingdom genetic networks. To decouple pathway evolution from the main genome and enable targeted, high-throughput genetic diversification, this study establishes an isolated neochromosome architecture. The methodology integrated hierarchical Level 1 Golden Gate assembly in Escherichia coli with Level 2 in vivo homologous recombination in yeast to construct functional five-gene human and seven-gene native yeast CoA neochromosomes. These constructs were flanked by orthogonal Dre-ROX recombination sites to permit independent structural modification, coupled with a multiplex CRISPR-Cas9 strategy to excise endogenous yeast loci and enforce metabolic reliance on the synthetic construct. Furthermore, a ratiometric fluorescent cytoplasmic CoA biosensor (PancACe) was integrated into a genomic safe-harbor locus to enable real-time metabolic monitoring. Results demonstrated the successful assembly and sequence-verified fidelity of these new synthetic chromosome constructs. Future uses of this platform extend to screening engineered strains for enhanced industrial production and serving as a versatile template for studying other complex human metabolic networks. The next steps for this project are to complete functional knockout testing to prove the synthetic chromosome can keep cells alive, finish validating the real-time sensor, and run gene-shuffling screens to isolate high-performing metabolic variants.

Keywords: Neochromosome, Coenzyme A, SCRaMbLE

Introduction

Coenzyme A (CoA) as a Critical Metabolic Hub

Coenzyme A (CoA) is an essential, universally ubiquitous metabolic cofactor required for approximately 4% of all known enzymatic transformations (Leonardi et al., 2005) . It serves as a foundational acyl carrier in cellular energetics, lipid biosynthesis, epigenetic histone modifications, and more (Barritt et al., 2024). Thus, in the eukaryotic domain, from the budding yeast *Saccharomyces cerevisiae* to humans, almost all organisms synthesize CoA *de novo* through the same five enzymatic steps from Pantothenate (Vitamin B5).

1. Phosphorylation: Pantothenate is phosphorylated by **pantothenate kinase (PANK)** to form 4'-phosphopantothenate.
2. Cysteine Addition: Cysteine is added to 4'-phosphopantothenate in an ATP-driven reaction catalyzed by **phosphopantothenoylcysteine synthetase (PPCS)** to yield 4′-phosphopantothenoylcysteine.
3. Decarboxylation: 4'-phosphopantothenoylcysteine is decarboxylated by **phosphopantothenoylcysteine decarboxylase (PPCDC)** to produce 4′-phosphopantetheine.
4. Adenylation: **Phosphopantetheine adenylyltransferase (PPAT)** transfers an adenyl group from ATP to form 3’-dephospho-CoA.
5. Phosphorylation: A final phosphorylation was carried out by **dephospho-CoA kinase (DPCK)**, which converts dephospho-CoA into active Coenzyme A using ATP. (Figure 1)

Dynamically probing the kinetic boundaries and evolutionary plasticity of this essential pathway is highly critical. It provides key insights into metabolic regulation and genetic cross-kingdom conservation.

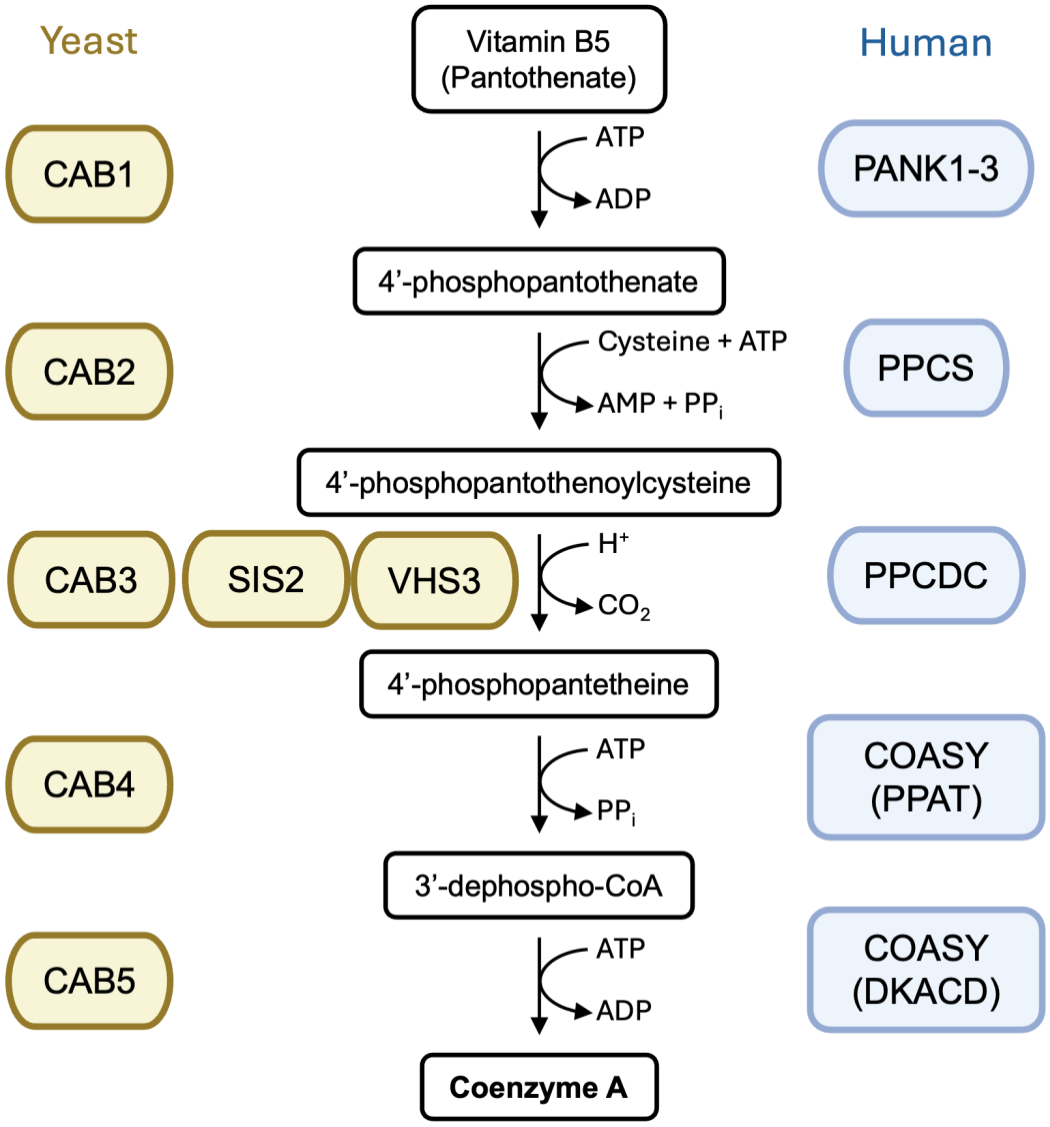

Figure 1. Coenzyme A Biosynthesis Pathway

Metabolic Engineering and SCRaMbLE

However, traditional metabolic engineering has its constraints. For instance, techniques like plasmid expression or regulator sequence swapping may ignore stoichiometric equilibrium that's particularly important in such complex pathways. Also, these techniques cannot generate and screen large amounts of genetic diversity simultaneously.

The Synthetic Yeast Genome Project (Sc2.0) has fundamentally redefined these boundaries. By introducing symmetrical loxPsym sites globally, it enables **SCRaMbLE** (Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution) (Dymond et al., 2011). The SCRaMbLE technique utilizes thousands of LoxPsym recombination sites downstream of genes to generate diverse genomic rearrangements, including deletions, inversions, duplications, and translocations (Swidah et al., 2020) These genomically diverse strains can then be evaluated under specific environmental stressors to identify optimal genome structures for metabolic engineering applications (Figure 2)..

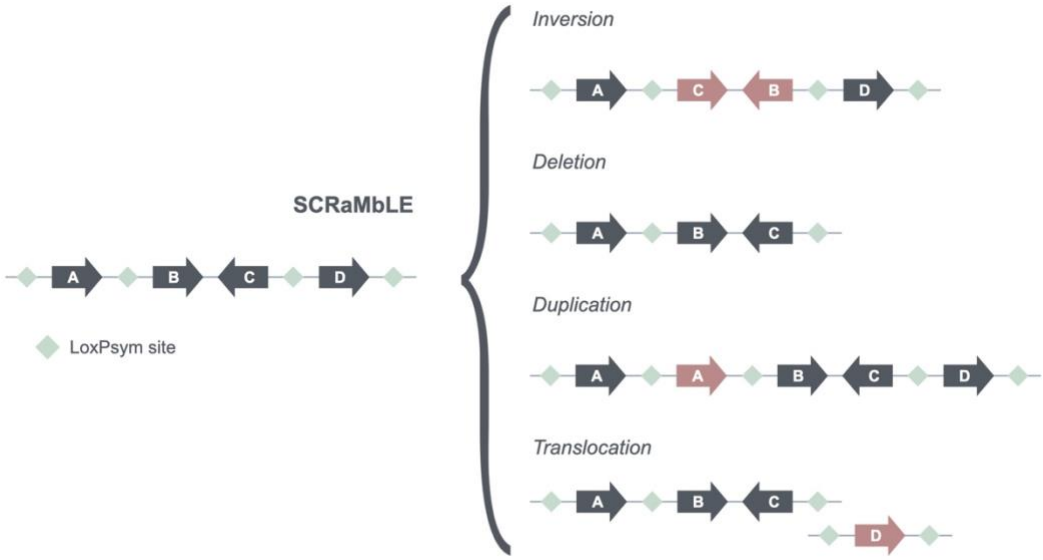

Figure 2. SCRaMbLE (Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution) Mechanism

While human genome scrambling has been reported, the survival rate of Cre-recombinase-incorporated strains has remained extremely low (Koeppel et al., 2025). To circumvent these barriers, an alternative approach would be to investigate the human genetic networks within the yeast environment, leveraging the fact that yeast scrambling has been well established in multiple previous studies. However, deploying this interspecies strategy effectively requires targeting a metabolic network that is both highly conserved and tightly regulated.

The Coenzyme A (CoA) biosynthesis pathway serves as an ideal model for this paradigm due to its exceptional evolutionary conservation. In humans, the five steps are governed by four different enzymes: PANK, PPCS, PPCDC, and COASY. The human enzyme **CoA synthase (COASY)** consists of two catalytic domains, PPAT and DPCK, that perform the last two steps simultaneously. On the other hand, the yeast genome utilizes seven different genes for these five steps: **CAB1**, **CAB2**, **CAB3**, **CAB4**, **CAB5**, along with **SIS2** and **VHS3**, which complement the CoA-synthesizing protein complex (CoA-SPC) (Figure 1).

Underscoring this functional conservation, functional cross-kingdom complementation has been demonstrated across all yeast genes. Namely, the CAB1 null mutant could be complemented by the human ortholog PANK3 (Ceccatelli Berti et al., 2022). (The human genome possesses three paralogs of the PANK gene; in this study we used the most complementable PANK3 paralog.) CAB2 was shown to be complementable by the human PPCS (Hamza et al., 2015; Iuso et al., 2018; Kachroo et al., 2015) CAB3, SIS2, and VHS3 null mutants can be rescued by human PPCDC (Ruiz et al., 2009); CAB4 can be complemented by human COASY (Hamza et al., 2015) ; and lastly, CAB5 can be complemented by human DCAKD (Dephospho-CoA Kinase Domain Containing) (Kachroo et al., 2015). This complete

compatibility makes it possible to study the whole humanized pathway together in a yeast background.

Neochromosome Engineering for Stable SCRaMbLE

While individual complementation studies pave the way for fully humanized models, standard SCRaMbLE systems are fundamentally constrained by low optimization efficiency. This bottleneck arises because roughly 20% of the yeast genome comprises essential genes (Z. Zhang & Ren, n.d.). Since scrambling introduces non-specific genomic deletions, random structural variations frequently trigger deletion of these crucial genes. The cells die upon deletion of any of these 1000 genes, hindering the identification of strains with optimal metabolic performance.

To overcome this structural constraint and mitigate host-lethal events, we propose a neochromosome architecture isolated from the main genome structure, using the CoA biosynthesis pathway as a proof of concept. We hypothesized that a synthetic neochromosome carrying the complete, five-gene mammalian Coenzyme A (CoA) pathway can entirely replace the endogenous yeast CoA machinery. This design enables the functional rescue of a multi-gene pathway knockout yeast strain, permitting the execution of SCRaMbLE within a stable yeast host to systematically investigate pathway plasticity. As a comparative control, we constructed an identical neochromosome harboring native yeast CoA genes to evaluate the functional impacts of chromosomal context and spatial localization.

Furthermore, we incorporated dual-layer genetic control by flanking the gene transcriptional units with Dre-ROX orthogonal recombination sites, operating independently of the background Sc2.0 Cre-loxP system. This secondary recombination layer allows for precise, independent tuning and structural modification of the CoA biosynthesis machinery. Finally, a

real-time acetyl-CoA reporter system was integrated to dynamically monitor and screen for high-performing metabolic strains. Taken together, this study represents the first example of multi-gene, cross-kingdom pathway integration that evaluates genetic orthogonality on a systematic level, establishing a novel framework for SCRaMbLE-mediated humanization in yeast.

Materials and Methods

Strains, Media and Growth Conditions

Bacteria

Escherichia coli strain DH5α was used in this study to clone Level 1 golden gate assembly products. Bacterial transformation was performed following the NEB standard high efficiency transformation protocol using chemically competent cells. *E. coli* transformants were cultivated in Lysogeny Broth (LB) liquid medium or on LB agar plates supplemented with 100 g/mL Ampicillin. Bacterial cultures were incubated at 37°C with shaking at 260 rpm overnight. For long-term storage, bacterial strains were preserved as 15% (v/v) glycerol stocks and stored at -80°C.

Yeast

Both the wild type *Saccharomyces cerevisiae* **BY4742** and the synthetic strain **Syn6.5** containing pre-integrated Cre-LoxP elements were utilized as the chassis for Level 2 neochromosomes in this study. Depending on experimental requirements, yeast strains were routinely grown in either standard YPD medium or Synthetic Complete (SC) selection medium with appropriate auxotrophic selection markers. Both media types were prepared as either liquid broth or solid agar plates. Yeast transformation was carried out using the standard lithium acetate (LiAc)/single-stranded carrier DNA/polyethylene glycol (PEG) method described Gietz and Schiestl (2007) (Gietz & Schiestl, 2007). Liquid cultures were incubated overnight at 30°C with shaking at 260 rpm. For solid media, agar plates were incubated at 30°C for 2–3 days until visible colonies appeared. Generally, the synthetic strain typically exhibits a slower growth rate

than the wild-type strain. For long-term torage, all yeast strains were preserved as 20% (v/v) glycerol stocks in their respective growth media and maintained at -80°C.

In Sillico Sequence Design

Sequence Retrieval and Codon Optimization

CoA biosynthesis coding sequences for the five human orthologs (PANK3, PPCS, PPCDC, DCAKD, COASY) were retrieved from the National Center for Biotechnology Information (NCBI) database. Correspondingly, the seven native yeast orthologs (CAB1, CAB2, CAB3, SIS2, VHS3, CAB4, CAB5) were mined from the Saccharomyces Genome Database (SGD).

To maximize translational efficiency and mitigate expression bottlenecks within the Saccharomyces cerevisiae chassis, all human-derived open reading frames (ORFs) were codon-optimized using the GeneScript Codon Optimization Tool, applying the host-specific S. cerevisiae codon usage frequency matrix. All sequences were computationally screened and modified in silico to eliminate internal Type IIS restriction sites, specifically BsaI and PaqCI. Computational design, vector layout visualization, and sequence annotations were managed using the Benchling and SnapGene platforms.

Hierarchical Level 1 and Level 2 Design Logic

Chimeric Level 1 (L1) transcription units (TUs) were designed using a standardized three-part expression cassette architecture. Optimized human or yeast ORFs were placed under the transcriptional control of native yeast promoters, defined as the 500 bp regions upstream of the native yeast start codon, and native yeast terminators, defined as the 200 bp region downstream of the native yeast stop codon (Figure 3). The start (ATG) and stop codons were strategically embedded directly within the flanking regulatory fragments to preserve intact

translational junctions. To facilitate Golden Gate assembly, specific Type IIS BsaI recognition sequences and standardized 4-bp fusion overhangs were appended to the boundaries of each part via PCR amplification using PrimeSTAR GXL DNA Polymerase (Takara Bio).

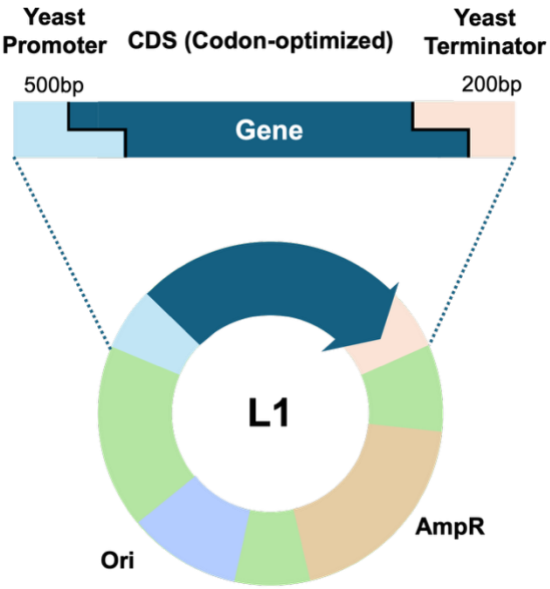

Figure 3. Level 1 Assembly Architecture

An exception was made for the yeast orthologs SIS2 and VHS3, which contain extensive aspartate and glutamate repeat regions that preclude complete commercial chemical synthesis. To circumvent this limitation, a hybrid assembly strategy was deployed: the repetitive terminal domain sequences were PCR-amplified directly from wild-type *S. cerevisiae* genomic DNA along with their native terminators, while only the non-repetitive upstream portions of the ORFs were codon-optimized and ordered as synthetic fragments. Furthermore, due to internal restriction sites, the SIS2 and VHS3 expression cassettes were engineered utilizing PaqCI restriction sites and customized overhangs instead of the standard BsaI framework.

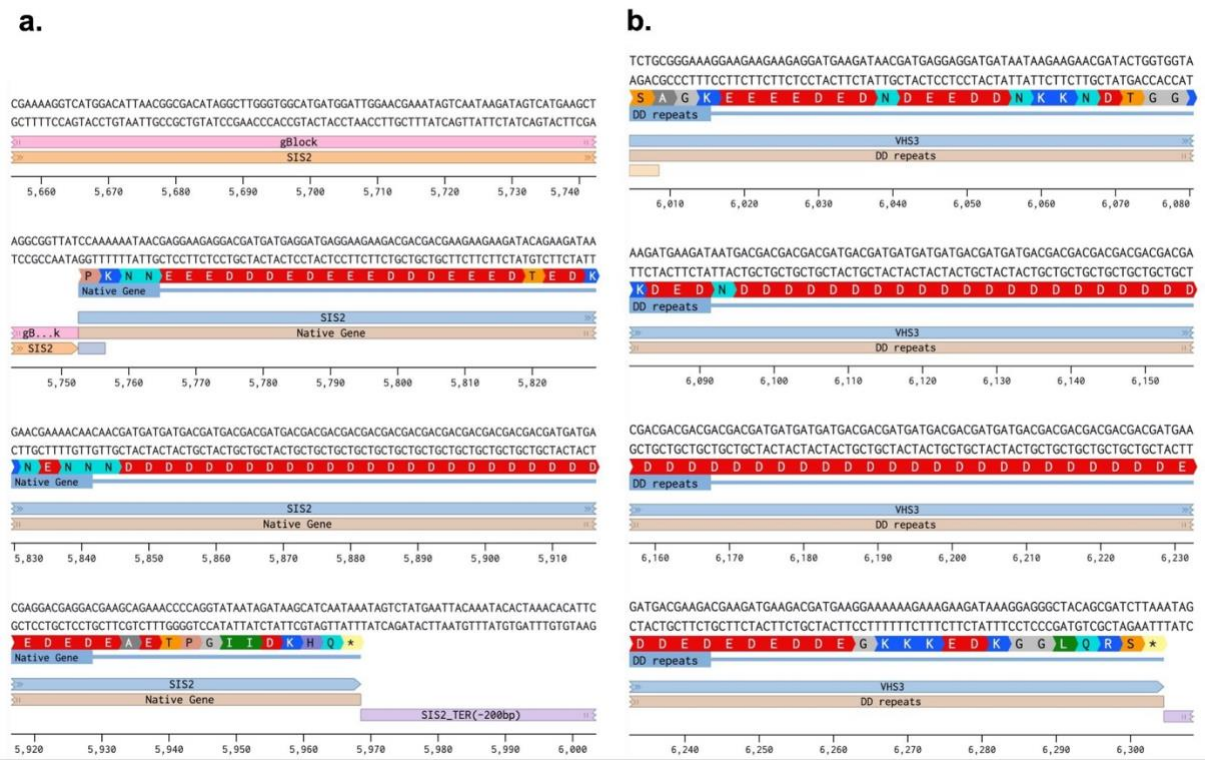

Figure 4. a. SIS2 C-terminus aspartate repeats; b. VHS3 C-terminus aspartate repeats

Level 2 (L2) neochromosomes were designed to concatenate individual L1 Transcriptional units into a single physical entity to facilitate yeast transformation. Unique, non-homologous Dre-ROX recombinase sites flanked with specific gRNA recognition sites were engineered as directional linkers between adjacent L1 units to allow for downstream genomic rearrangement. (Figure 5)

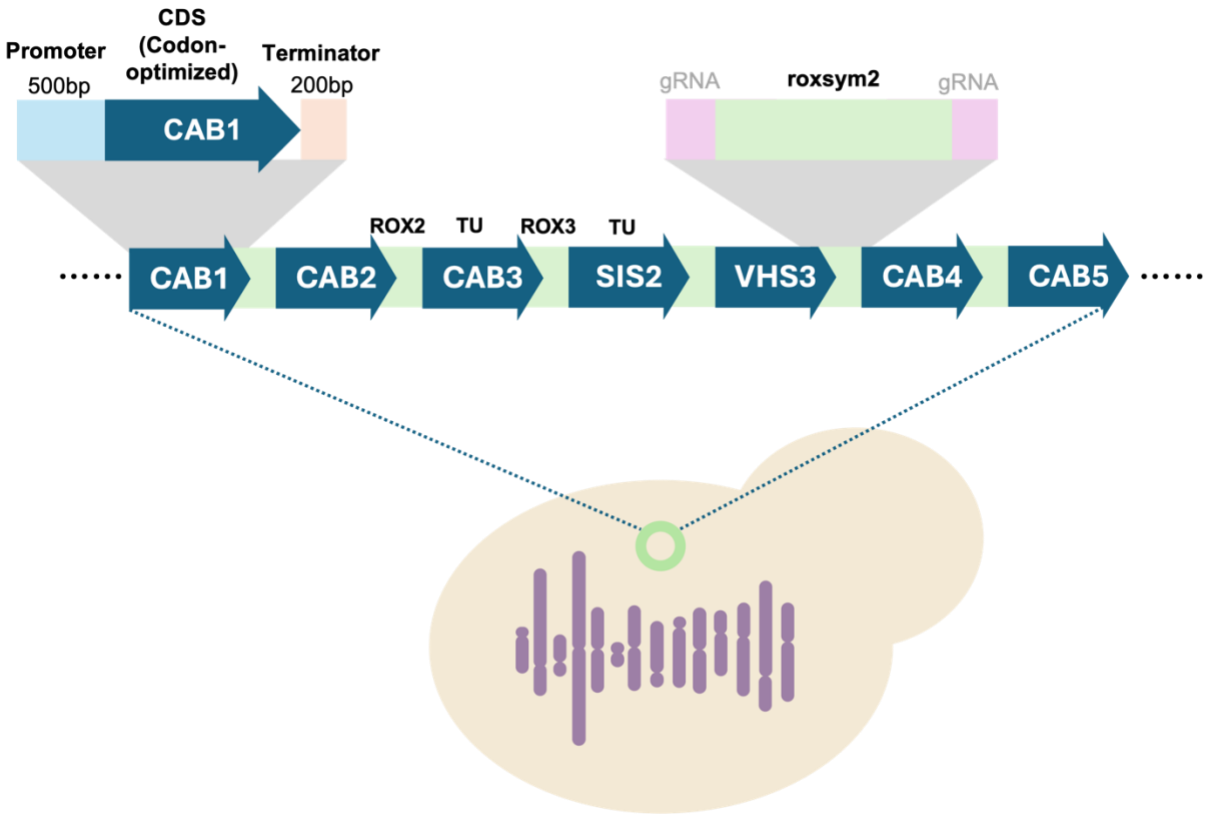

Figure 5. Level 2 (L2) Design Architecture

Hierarchical In Vitro and In Vivo Assembly of Neochromosomes

Level 1 Transcription Unit Assembly (PCR, Golden Gate)

Synthetic human and yeast ORFs were synthesized after codon optimization by Integrated DNA Technologies (IDT). Native promoters and terminators were amplified from wild-type yeast strain BY4742 genomic DNA using PrimeSTAR GXL DNA Polymerase (Takara Bio).

Level 1 transcriptional units were assembled via bacterial Golden Gate after PCR purification using the PureLink™ PCR Purification Kit (Invitrogen). Each reaction was performed in 20µL solution containing 25 fmol each of the respective promoter, ORF, and terminator fragments, 25 fmol of the L1 destination vector, 1 µL T4 Ligase, and 1 µL BsaI in 1× T4 DNA Ligase Buffer. The thermocycler profile consisted of an initial digestion at 37°C for 5

min, followed by 30 cycles of digestion at 37°C for 1 min and ligation at 16°C for 1 min. Then ending with a final ligation at 16°C for 10 min, final digestion at 37°C for 20 min, and heat inactivation at 65°C for 10 min.

An exception was made for the SIS2 and VHS3 transcription unit assemblies. These reactions were performed in a 20 µL final volume containing 25 fmol each of the respective promoter, ORF, and terminator fragments, alongside 25 fmol of the PCR-amplified L1 destination vector modified to include PaqCI restriction sites. The assembly mixture was supplemented with 1 µL PaqCI enzyme, 0.5 µL PaqCI Activator, and 1 µL T4 DNA Ligase in 1× T4 DNA Ligase Buffer. The thermocycler profile was the same as for BsaI assemblies.

Level 1 Bacterial Transformation and Screening

The L1 reaction products were transformed into competent *E. coli* DH5α cells via heat-shock at 42°C for 45 s, recovered in Lysogeny Broth (LB) medium for 1 h at 37°C, and selection-plated onto LB agar containing 100 µg/mL ampicillin and Blue-White Screening Reagent (Merck) following the NEB standard high-efficiency transformation protocol. Plates were incubated overnight at 37°C, and successful transformants were isolated via blue-white colony screening. White colonies were cultured overnight in liquid selective LB broth, and plasmids were extracted using a E.Z.N.A.® Plasmid DNA Mini Kit (Omega bio-tek). The complete sequence fidelity and structural integrity of all L1 plasmids were verified via Next-Generation Sequencing (NGS) provided by Plasmidsaurus.

Level 2 Neochromosome Assembly (Homologous Repair)

To prepare the building blocks for in vivo assembly, individual transcription units (either the 5 human ortholog cassettes or the 7 yeast ortholog cassettes) were PCR-amplified to introduce unique, overlapping homologous sequences between adjacent fragments. The

neochromosomes were constructed in the following order: Human neochromosome: PANK3, PPCS, PPCDC, COASY, DCAKD; Yeast Neochromosome: CAB1, CAB2, CAB3, SIS2, VHS3, CAB4, CAB5. Concurrently, the yeast centromeric shuttle vector pRSII413_iCEN8 was linearized and functionalized with terminal homology arms matching the outermost boundaries of the gene arrays. All PCR products were purified to remove residual templates and primers.

The purified, linear DNA fragments were pooled into two distinct multiplex assembly mixtures: a 6-fragment pool for the humanized neochromosome (5 human cassettes plus the linearized pRSII413_iCEN8 backbone) and an 8-fragment pool for the yeast control neochromosome (7 yeast cassettes plus the backbone). These multi-fragment pools were simultaneously co-transformed into competent *S. cerevisiae* cells (BY4742 and Syn6.5) using a high-efficiency LiAc protocol described by (Gietz & Schiestl, 2007). Intracellular assembly was driven entirely by the endogenous homologous recombination machinery of the yeast hosts.

Level 2 Systematic Screening and Verification

Transformants were selected on Synthetic Complete media lacking histidine (SC–His-Leu) to ensure the retention of the functionalized pRSII413_iCEN8 backbone (-His) and the Cas9 plasmid (-Leu).

To verify full-length circularization and accurate multi-fragment alignment, surviving colonies were screened via multiplex junction colony PCR following either the GoTaq or the DreamTaq protocol. Primer pairs were strategically designed to span every adjacent gene-to-gene and gene-to-backbone interface. Successful assemblies were identified by the presence of distinct, diagnostic bands of the expected amplicon size on a 1% agarose gel. Further verification was performed by sequencing the PCR amplicons via NGS of Plasmidsaurus. For long-term

archiving, all validated yeast strains were preserved as 20% (v/v) glycerol stocks and maintained at -80°C.

Biosensor Reporter Integration

Reporter Constructs

The ratiometric fluorescent cytoplasmic CoA sensor PancACe and its negative control cpGFP (circularly-permuted GFP) were built in this study as the in vivo biosensor (Smith et al., 2025). In accordance with the hierarchical design rules established above, the synthesized open reading frames (ORFs) were codon-optimized via GenScript and assembled into Level 1 destination vectors to be constitutively driven by the yeast TEF1 promoter (PTEF1) and terminated by the CYC1 terminator (TCYC1). Bacterial transformation, blue-white colony screening, plasmid extraction, and Plasmidsaurus NGS verification were carried out identically to the standard Level 1 protocol described above.

CRISPR-Cas9 Mediated Integration

To achieve stable genomic integration, the XI-5 locus on yeast chromosome XI was targeted as a genomic safe harbor. The sequence-verified PTEF1-PancACe-TCYC1 and PTEF1-cpGFP-TCYC1 cassettes were PCR-amplified using PrimeSTAR GXL DNA Polymerase to introduce 40-bp terminal homology arms directly matching the flanking regions adjacent to the targeted safe-harbor insertion site. The linear fragments were co-transformed with the Cas9 expression vector and the gRNA plasmid mapping the locus using the same three-component CRISPR editing system. Successful integration was validated via colony PCR using verification primers targeting genomic regions external to the original homology boundaries.

Multiplex CRISPR-Cas9 Endogenous Gene Deletion

To assess functional complementation and force cellular dependence on the newly introduced L2 synthetic neochromosomes, a 3-component multiplex CRISPR-Cas9 editing system was designed to simultaneously disrupt all endogenous pathways. The architecture consisted of a localized Cas9 expression vector (with a Leucine auxotrophic marker), a single multiplex gRNA vector targeting all vital endogenous pathways, and marker-free linear donor repair templates. The donor templates were engineered via the same Golden Gate assembly process to fuse the 500 bp upstream flanking region directly to the 500 bp downstream flanking region of each native gene. Donor templates were introduced to enforce clean, scarless gene erasure via Homology-Directed Repair (HDR) in yeast. Functional complementation survival was benchmarked by screening for viable strains post-transformation.

Results

1. Adaptation and Integration of a Genetically Encoded Cytoplasmic CoA Biosensor

To screen genetically modified yeast strains after SCRaMbLE, we needed a way to quantify Coenzyme A (CoA) production without destroying cells or disrupting normal biological functions. As described in previous works, overexpression of CAB1-CAB5 can lead to an increase in CoA production and lead to a massive accumulation of acetyl-CoA (Olzhausen et al., 2021). Thus, in this research, we focus on measuring Acetyl-CoA to indicate the activity of our CoA biosynthesis neochromosome and rely on real-time acetyl-CoA levels to monitor SCRaMbLE strain activities.

While mass spectrometry and assay kits can measure acetyl-CoA in cell lysates, these approaches cannot be used in living cells in real time. Initially, an engineered heterologous β-carotene reporter pathway was integrated into the yeast genomic XI-5 safe harbor to track metabolic activity colormetrically (López et al., 2020). However, this multi-step downstream reporter imposed a severe metabolic burden and provided an indirect readout that did not correlate linearly with real-time cytoplasmic CoA pools.

To overcome these limitations, we adapted PancACe, a genetically encoded, ratiometric fluorescent biosensor previously validated in bacterial and mammalian systems, for deployment in *S. cerevisiae* (Smith et al., 2025). The genetically encoded fluorescent biosensor **PancACe** consists of two components: a circularly permuted GFP (cpGFP) that provides the fluorescence and the PanZ domain that detects CoA.

As this sensor had never been deployed in yeast, we codon-optimized and synthesized both PancACe and its negative control cpGFP as gBlocks. Then cloned them downstream of the

strong constitutive TEF1 promoter and upstream of the CYC1 terminator in *E. coli* (Figure 6a). Following sequence verification, the PancACe and cpGFP expression cassettes were PCR-amplified and integrated into the yeast genomic XI-5 locus (a known genomic safe harbor from López et al. (2020) via CRISPR-Cas9-mediated homologous recombination. Integration was validated via colony PCR and sequencing (Figure 6b). Although full functional fluorescent validation remains an ongoing laboratory objective due to the project's tight two-month timeframe, the genetic platform for real-time CoA monitoring was successfully established in vivo.

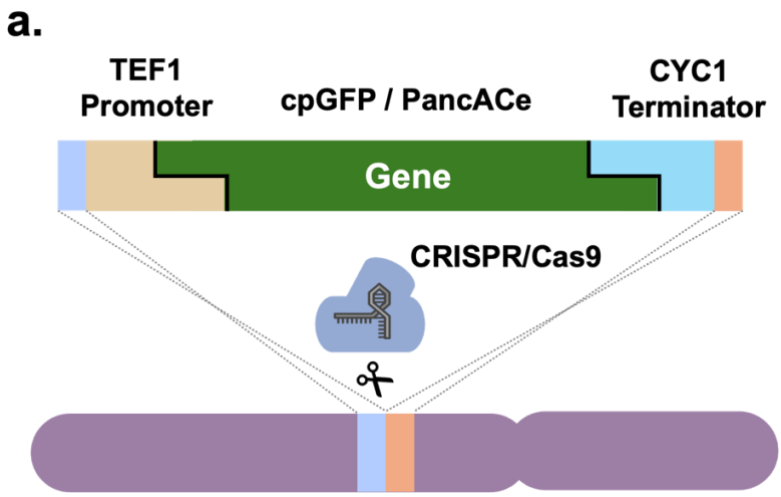

Figure 6a. Genetic Architecture of cpGFP/PancACe cassette

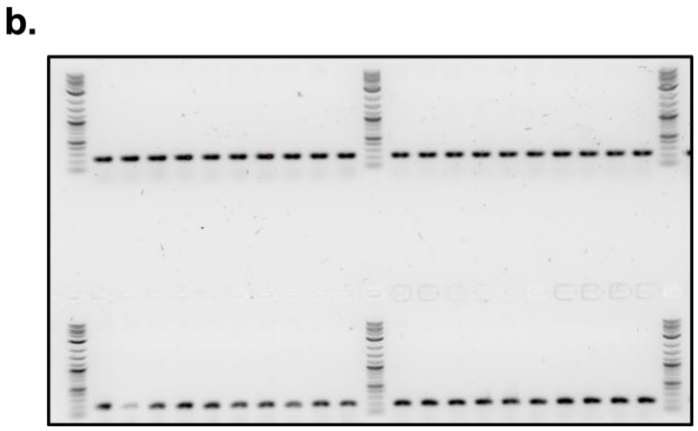

Figure 6b. Colony PCR verification (n=80). All colonies screened failed integration.

2. Multi-Tiered Design and Hierarchical Assembly of Yeast and Human CoA Neochromosomes

To evaluate the cross-species orthogonality and functional boundaries of human CoA biosynthesis enzymes in a yeast host, we engineered two distinct artificial chromosomes consisting of the CoA Biosynthesis pathway of human and the native yeast control group and expressed them in *S. cerevisiae*.

Design Strategy and Level 1 (L1) Assembly

Because human gene regulatory sequences are non-functional in *S. cerevisiae*, engineering the neochromosome required cloning human genes into native yeast promoter and terminator structures. Five human orthologs (PANK3, PPCS, PPCDC, COASY, and DCAKD) and seven native yeast orthologs (CAB1, CAB2, CAB3, CAB4, CAB5, SIS2, VHS3) were codon-optimized, synthesized, and cloned downstream of native yeast promoters (~500 bp) and upstream of terminators (~200 bp). Specifically, PANK3 was flanked by the CAB1 promoter and terminator sequence, PPCS was flanked by the CAB2 promoter and terminator sequence, and so on to CAB3, CAB4, and CAB5 (Figure 7a). The seven yeast ortholog CDSs were also codon-optimized to enable CRISPR deletion to only delete the native copies of each yeast gene. All twelve distinct L1 transcription units were successfully assembled in *E. coli*, screened with ampicillin, and sequence-verified (Figure 7b).

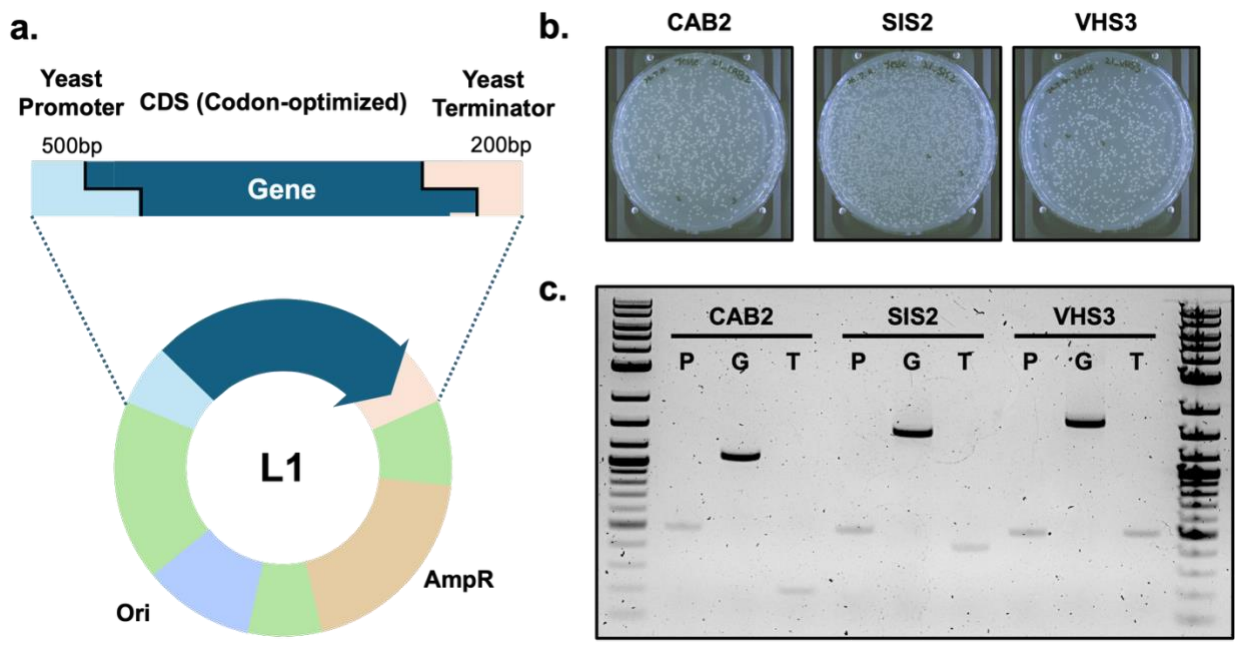

Figure 7. a. L1 Assembly Architecture; b. Successful L1 colonies (using CAB2, SIS2, VHS3 as example) c. Gel confirmation of L1 assembly fragments (using CAB2, SIS2, VHS3 as example); NEB 1kb plus ladder P: Promoter PCR sequence (~500bp), G: Gene PCR sequence (~1kb), T: Terminator PCR sequence (~200bp)

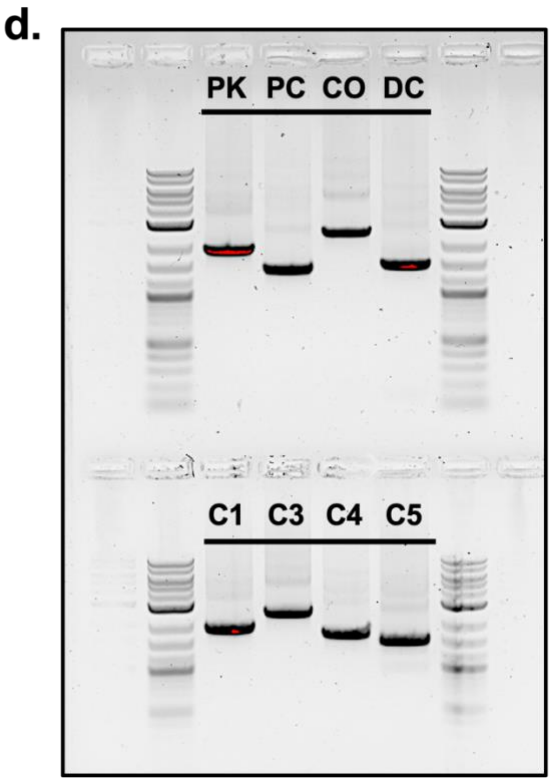

Figure 7. d. Gel confirmation of L1 Transcriptional Units (choose 8 for example: PK: PANK3, PC: PPCDC, CO: COASY, DC: DCAKD, C1: CAB1, C2: CAB3, C4: CAB4, C5: CAB5); using NEB 1kb plus ladder

Level 2 (L2) Neochromosome Assembly and Yeast Integration

Once we acquired Level 1 transcriptional units of all twelve genes consisting of a yeast promoter followed by a codon-optimized CDS, then ending with native yeast terminators, we moved on to assembling L2 neochromosomes. Both neochromosomes were built from the pRSII413_iCEN8 backbone consisting of CEN/ARS replication sites, HIS3 auxotrophic markers, and homology arms (Figure 8a). The neochromosome was built utilizing the in vivo homologous repair system in yeast by adding homologous arms on the two ends of transcriptional units via PCR amplification.

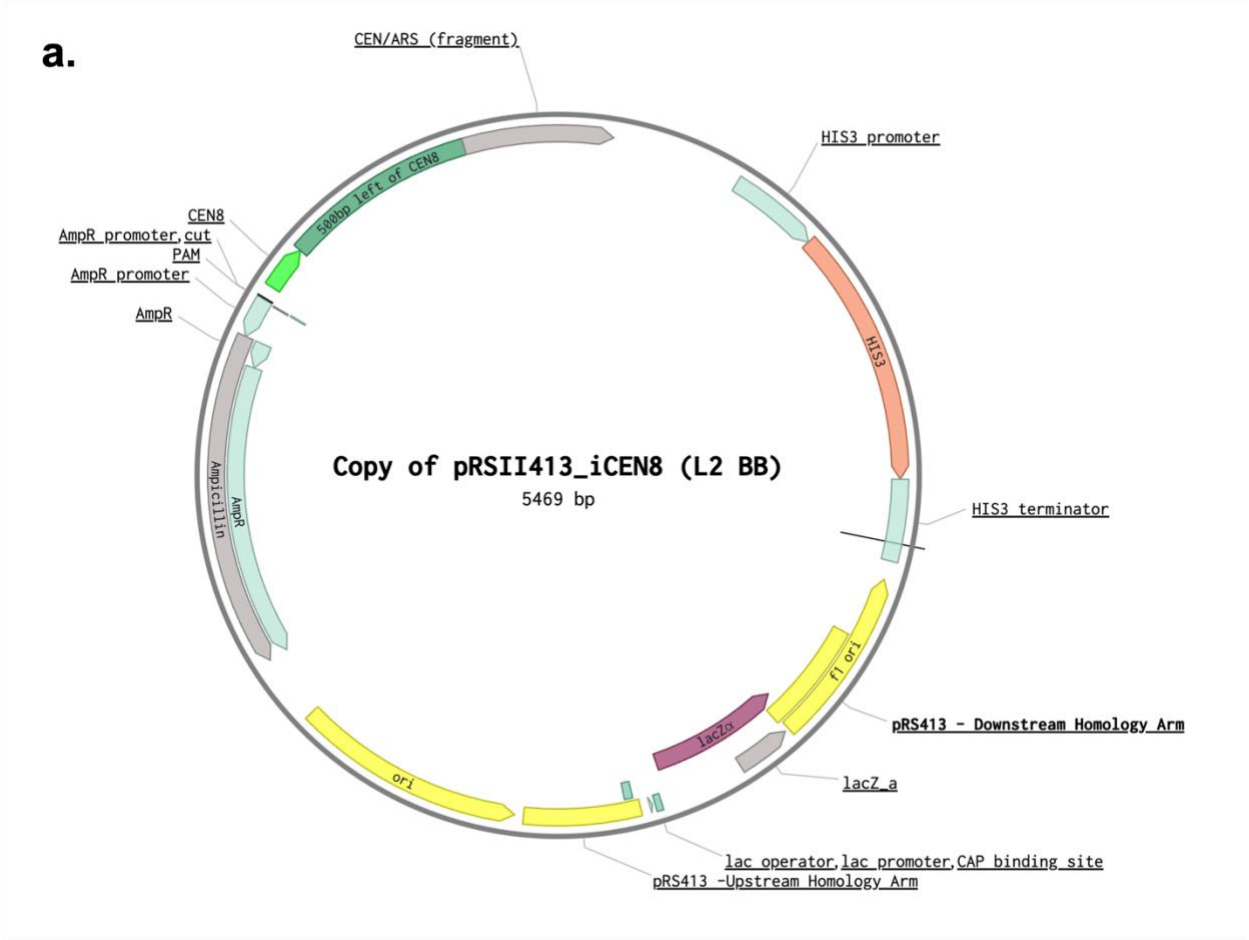

*Figure 8*a. Level 2 Destination vector

The human neochromosome consisted of five TUs of human orthologs following their biochemical function order: PANK3, PPCS, PPCDC, COASY, and DCAKD (Figure 8b). On the other hand, the yeast neochromosome consisted of the seven yeast genes in the CoA synthesis pathway assembled in its biochemical pathway order: CAB1, CAB2, CAB3, SIS2, VHS3, CAB4, CAB5 (Figure 8b). Each transcriptional unit was flanked with ROX linkers comprising two unique gRNA recognition sites flanking a specific ROXsym sequence (Figure 8c). The Dre-ROX system was chosen to separate the neochromosome SCRaMbLE process from that of the synthetic strain, which uses the Cre-LoxP system. Among the Dre-ROX sequences in Wang et. al (2022) research, we chose roxsym2 based on its symmetric ability that grants equal possibilities

for swapping genes and simultaneously having the highest inversion rate to perform efficient SCRaMbLE (Wang et al., 2022).

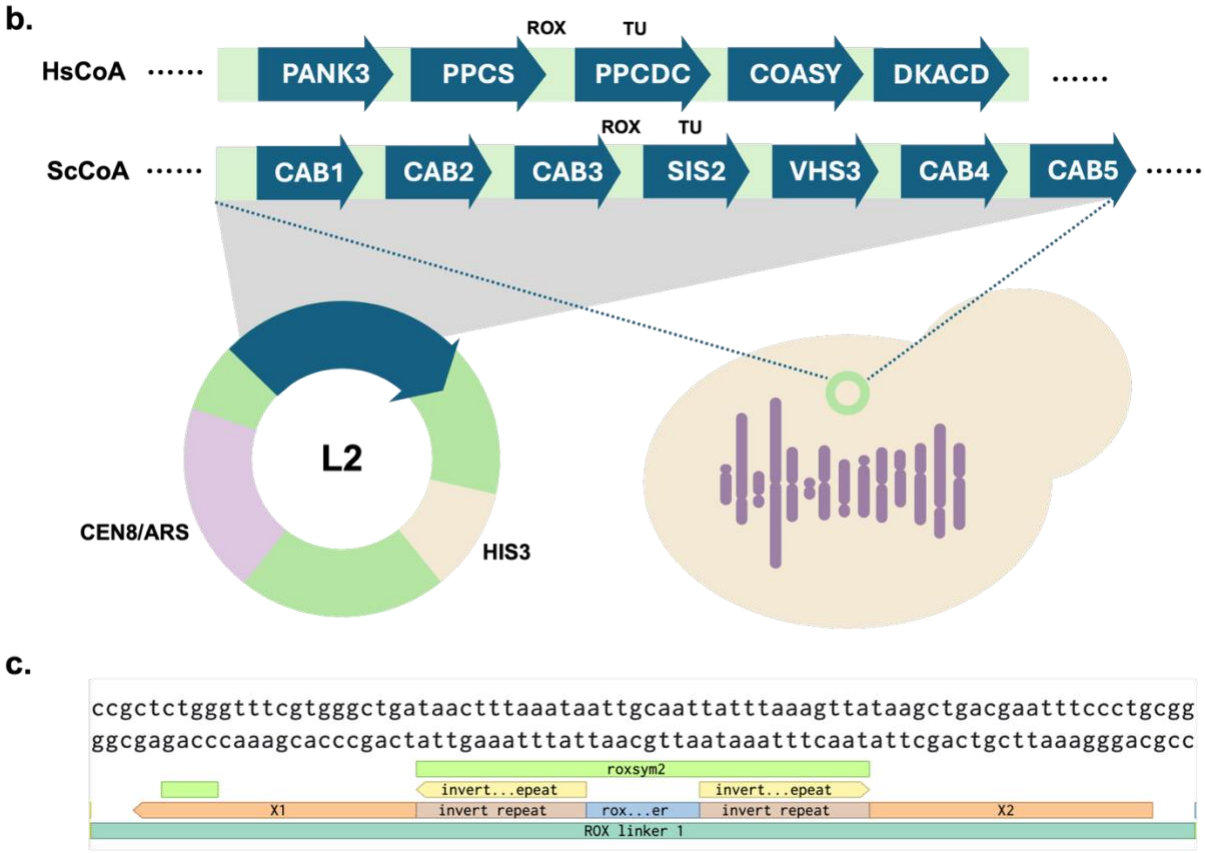

Figure 8b. L2 HsCoA and ScCoA design architecture; *8c*. ROX linker gene structure

Each of the ROX linkers flanking the transcriptional units is unique. However, they share the same roxsym2 sequence. Based on different gRNA sites flanking the sequence, they are named ROX linker 1-7. Here we use the linker between the right homology arm of the backbone and the CAB1 transcription unit, ROX linker 1, as an example.

L2 neochromosomes were assembled from co-transformation of each of the transcriptional units and the pRSII413_iCEN8 L2 backbone, PCR-amplified, including homology arms, using the LiAc transformation protocol (Gietz & Schiestl, 2007).They were then

transformed into wild-type (BY4742) and synthetic yeast strain Syn6.5 to compare their behavior. Clean, expected band sizes from multiplex junction colony PCR combined with long-read sequencing confirmed error-free, scarless neochromosome assembly and stable episomal maintenance.

3. CRISPR-Cas9 Multiplex Deletion of Endogenous Yeast CoA Genes

To investigate the functionality and evolutionary boundaries of the newly introduced neochromosome, the native yeast gene counterparts have to be evacuated. By knocking out all seven of the native yeast genes, we force the yeast strains to rely solely on the synthetic neochromosome. We executed a synchronized, seven-locus multiplex CRISPR-Cas9 deletion targeting all native yeast genes simultaneously: CAB1, CAB2, CAB3, SIS2, VHS3, CAB4, and CAB5. Crucially, since the gene ORFs consisting of the yeast neochromosome underwent codon optimization, their sequence diverged significantly from the original loci. This difference ensured that the designed gRNA would exclusively target the native genes and leave the introduced neochromosome intact.

To achieve this, a three-component CRISPR-Cas9 genome editing system was deployed. The editing machinery was delivered via a constitutively expressed Cas9 endonuclease vector bearing a LEU2 auxotrophic selection marker, co-transformed alongside a single multiplex gRNA expression vector.

Simultaneous cleavage of all seven endogenous loci was achieved by deploying a polycistronic gRNA-tRNA array architecture (Y. Zhang et al., 2019), which leverages host endogenous RNase P and RNase Z processing enzymes to internally cleave and release each distinct targeting sequence. Concurrently, homologous repair was directed by seven independent donor DNA templates engineered via Golden Gate assembly in *E. Coli*. By directly fusing the

~500 bp upstream and ~500 bp downstream flanking regions of each of the seven native targets, these amplicons facilitated a completely markerless and scarless excision of the genomic ORFs upon Cas9 cleavage (Figure 9).

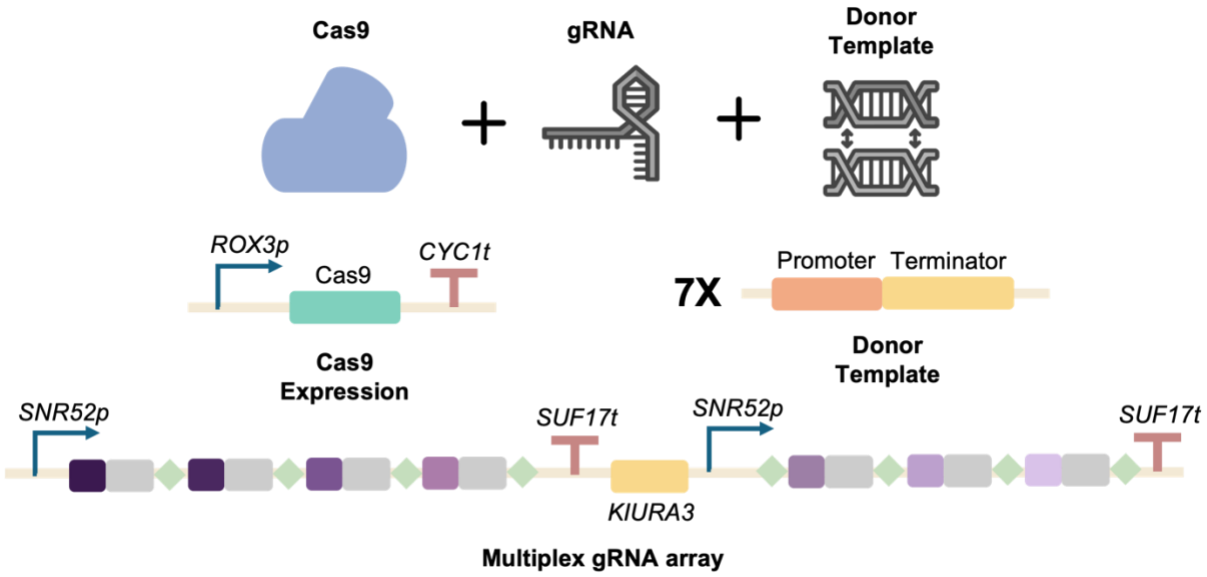

Figure 9. CRISPR-Cas9 Multiplex Deletion

Discussion

Technical Pivoting and Troubleshooting the In Vivo PancACe Biosensor Integration

To effectively screen strains immediately following SCRaMbLE, deploying a live system to measure metabolic output without destroying the cells is a critical objective of this study. In this study, our initial engineering iteration leveraged an integrated, heterologous β-carotene synthesis gene array at the genomic XI-5 safe harbor locus (López et al., 2020). By introducing three plant genes that convert IPP and DMAPP, both downstream products of acetyl-CoA, into β-carotene (Figure 4A), we are able to visualize and quantify acetyl-CoA production. While this colorimetric pathway successfully generated visible orange colonies, indicating a functional safe-harbor expression, deep operational limitations emerged.

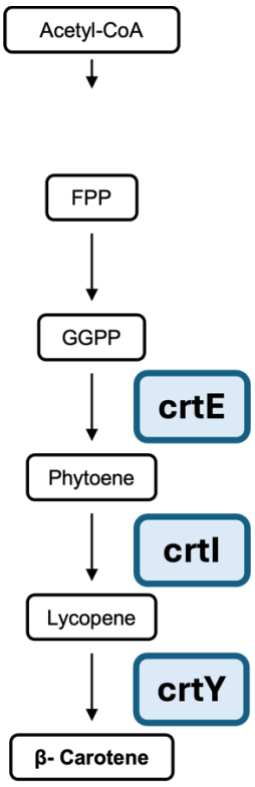

Figure 10. β-carotene pathway (Blue genes: crtE, crtI, and crtY are extrinsic genes introduced)

First, the heterologous production of carotenoids creates a significant metabolic sink. By pulling essential downstream metabolites (such as IPP and DMAPP) away from vital endogenous metabolic networks, this configuration artificially decreases host fitness. Because downstream SCRaMbLE screens rely heavily on cellular survival and fitness under severe environmental stress, utilizing a screening tool that inherently drains the cell's energetic reserves would undermine the selective pressure of the experiment.

Second, the long, multi-step enzymatic distance between the immediate synthesis of Acetyl-CoA/CoA and the final formation of β-carotene introduces substantial signal delay and non-linear kinetics. This structural separation makes it impossible to confidently conclude that a higher colorimetric intensity directly correlates with superior upstream neochromosome output.

To establish a direct and non-destructive screening pipeline, this study pioneered the porting and adaptation of the ratiometric fluorescent biosensor PancACe into *S. cerevisiae*. Although Golden Gate cloning in *E. coli* was structurally verified via NGS, successful genomic integration at the XI-5 locus could not be achieved, consequently preventing any functional fluorescent readouts within our operational timeframe. We hypothesize that two primary engineering bottlenecks contributed to this performance limitation:

1. Locus-Specific Chromatin Position Effects: While the XI-5 locus is a widely utilized genomic safe harbor for standard expression cassettes, eukaryotic integration sites are highly sensitive to their surrounding genomic architecture. Budding yeast features a highly compact genome where open reading frames (ORFs) occupy roughly 75% of the total sequence (Goffeau et al., 1996). Consequently, transgenes driven by highly active, constitutive promoters (such as TEF1) are frequently subject to transcriptional read-through from neighboring

native loci (Boldogköi, 2012) or localized nucleosome occlusion (Lee et al., 2007). This local chromatin environment can alter transgene expression profiles or dramatically compress the dynamic range of integrated ratiometric biosensors, necessitating the downstream use of structural insulators to shield the sensor platform.

2. Structural Anomalies within the Transfection Vector Background: Issues with the Cas9 plasmid backbone likely slowed things down and limited protein production. Although the vector passed our junction PCR checks, sequencing revealed potential structural deletions or truncations within the Cas9 sequence itself. Even minor flaws in this framework can disrupt nuclear localization, protein stability, or guide RNA loading. Ultimately bottlenecking overall performance and preventing detectable sensor signals within our experimental timeframe.

Notable Features of the Neochromosome Design and Structural Insight of Hierarchical Neochromosome Assembly

The genomic architecture of the HsCoA and ScCoA neochromosomes introduces several distinct structural advantages that optimize cloning workflows, ensure metabolic stability, and establish an advanced platform for pathway evolution. By transitioning from a traditional linear plasmid architecture to a specialized, episomal neochromosome, this design incorporates five critical features that collectively define a novel standard for macro-scale metabolic engineering.

1. Dual-Host Hierarchical Assembly Logic

Assembling multi-gene pathways directly in a Golden Gate single-pot reaction often causes a sharp exponential drop in assembly efficiency as fragment numbers increase. To bypass

this bottleneck, we implemented a dual-host hierarchical assembly strategy that exploits the unique genetic strengths of two distinct model organisms:

- Level 1 (L1) Assembly in *E. coli*: Individual transcriptional units (TUs) were cloned using highly predictable Golden Gate assembly methods. This bacterial phase facilitates faster generation turnovers and convenient sequence verification, allowing errors to be quickly isolated and debugged at the single-gene level.
- Level 2 (L2) Assembly in *S. cerevisiae*: Once verified, the L1 cassettes were captured into complete neochromosomes using yeast's highly efficient in vivo homologous recombination machinery. Homologous sequences were attached to the ends via PCR amplification necessary for assembly. Utilizing the host cell as the final cloning destination eliminates low-yielding in vitro macro-ligations and accelerates chromosome construction.

2. Native Eukaryotic Architecture for Humanized Orthologs

A major hurdle in cross-species transplantation is that heterologous mammalian enzymes wouldn't operate efficiently within a yeast cellular environment. Because human gene regulatory sequences would not be expressed in *S. cerevisiae*, we designed a native transcriptional framework for each humanized gene. Flanking each codon-optimized human open reading frame (ORF) with a native yeast promoter (~500 bp) and terminator (~200 bp) from its corresponding ortholog establishes predictable structural boundaries and spacers between transcriptional units. This architecture provides balanced, native-like transcription levels that ensure smooth metabolic integration without causing host toxicity. Crucially, this approach maintains the precise stoichiometric expression required for these metabolic pathways.

3. Symmetrical Linker Arrays and Orthogonal SCRaMbLE Systems

To enable downstream structural diversification, each transcriptional unit in the level 2 neochromosome is flanked by highly unique ROX linkers incorporating a symmetrical roxsym2 sequence (Wang et al., 2022). This spatial design unlocks several advanced analytical capabilities:

- Pathway Topology Mapping: The symmetrical linker design balances the probabilities of gene inversion, duplication, and excision among different genes. Inducing rearrangements uniformly between different genes allows us to study the regional effects of gene position and directly investigate whether clustering functionally related genes enhances overall pathway activity.
- Dual-Layer Chromosomal Modification: By utilizing the Dre-ROX system on the neochromosome, we successfully freed the neochromosome SCRaMbLE system from the host genome's existing Cre-LoxP background. This orthogonal setup grants researchers full combinatorial control: they can selectively mobilize the neochromosome path, mutate the host background, or trigger both systems simultaneously to evaluate indirect genomic noise on Coenzyme A synthesis.

4. Synthetic Recoding as a Genetic Shield

To completely inactivate native host genes, the open reading frames on the neochromosomes were synthetically recoded and codon-optimized. This modification introduces nucleotide sequence divergence from the endogenous genomic loci while fully conserving the essential amino acid configurations. This deliberate sequence acts as a "genomic shield" that allows multiplex CRISPR-Cas9 guide RNAs to selectively target and destroy all native genomic

genes simultaneously without causing off-target collateral damage to the newly introduced neochromosomes.

5. Topological Isolation of the Essential CoA Pathway

Ultimately, because Coenzyme A is an essential housekeeping molecule, direct genomic engineering of its pathway is heavily restricted by natural selection, limiting direct genomic engineering. By insulating the Coenzyme A pathway from the rest of the genome, scientists can explore the genome without worrying that it might affect its normal physiological activities. This design packages the humanized CoA pathway into a self-contained, modular unit. By separating it from the rest of the cell, researchers can safely modify and test essential metabolic networks across very different species without risking cell death.

Concept Framework for Dual-System SCRaMbLE and High-Throughput Screening Evolution

The integration of the Coenzyme A (CoA) pathway onto an independent Level 2 neochromosome provides downstream researchers with a powerful, tunable platform for metabolic engineering. By utilizing the orthogonal Dre-rox and Cre-loxP recombinase systems, scientists can actively drive genomic diversity, screen for high-performing phenotypes, and dissect complex host-pathway interactions. This section outlines the chronological pipeline for leveraging this architecture to optimize CoA production in *Saccharomyces cerevisiae*.

1. Generating Genomic Diversity via Combinatorial SCRaMbLE

The dual-recombinase design allows researchers to direct structural rearrangements precisely, offering three distinct options to introduce genomic diversity:

1. Dre-mediated SCRaMbLE: This approach acts exclusively on the roxsym2-flanked human pathway genes on the neochromosome. It allows researchers to

systematically investigate how different gene orders, orientations, and physical positions affect CoA production. Furthermore, it serves as a functional tool to identify rate-limiting steps or determine whether specific pathway configurations alter feedback regulation.

2. Cre-mediated SCRaMbLE: This approach targets the endogenous loxP sites distributed throughout the synthetic host genome (Syn6.5 background) while leaving the humanized neochromosome intact. It allows researchers to explore how distal host genes, global regulatory networks, and native metabolic pathways influence human CoA synthesis.
3. Simultaneous Dre + Cre SCRaMbLE: By activating both recombinases, the host genome and the heterologous pathway are forced to co-evolve. This dual-scrambling setup enables the discovery of complex synergistic effects or reveals whether global down-regulation events in the host override local pathway improvements or vice versa.

While all three configurations generate substantial genetic diversity, we predict that Cre-only and joint Dre + Cre SCRaMbLE will yield the most biologically interesting and unexpected results. Because the five human genes (PANK3, PPCS, PPCDC, COASY, DCAKD) operate sequentially in a strict linear cascade, random Dre-mediated deletions on the neochromosome are highly likely to break the pathway entirely, resulting in loss-of-function or lethal phenotypes. Similarly, simple duplications of a rate-limiting step are expected to yield predictable, incremental increases in yield.

In contrast, scrambling the host genome via Cre or Dre + Cre opens an unbiased, genome-wide search. This strategy is uniquely capable of uncovering unexpected helper genes,

novel transcription factors, or distal metabolic nodes that were not previously known to participate in or regulate the CoA pathway.

2. Two-Phase High-Throughput Screening Sieve

To efficiently isolate high-producing strains from deep combinatorial libraries, we propose a two-phase selection framework integrated with the PancACe biosensor platform:

- Phase 1 (The Glycerol Sieve): SCRaMbLEd library variants undergo primary screening via culture on non-fermentable glycerol media compared to standard glucose control plates. Because CoA and Acetyl-CoA are absolutely essential for aerobic respiration and the tricarboxylic acid (TCA) cycle, yeast cannot metabolize glycerol anaerobically. Broken, sluggish, or poorly performing pathway variants will fail to respire and will be eliminated, serving as a robust survival filter.
- Phase 2 (PancACe Biosensor Validation): Surviving strains from the primary sieve are monitored in real time using the integrated ratiometric PancACe fluorescent biosensor. While the reporter system was not fully operational within the immediate timeframe of this study, its architectural implementation allows researchers to directly quantify intracellular Acetyl-CoA concentrations. High-performing variants can be rapidly sorted using microplate readers for population assays or single-cell Fluorescence-Activated Cell Sorting (FACS) for ultimate high-throughput efficiency.

To continuously stack beneficial genomic modifications and maximize production phenotypes, strains should be subjected to iterative rounds of this pipeline, alternating between phases of SCRaMbLE induction and rigorous screening.

3. Downstream Genomic Mapping

After finding the best-performing strain, Nanopore long-read sequencing could be applied to sequence the whole genome. Because SCRaMbLE causes massive structural shuffling (such as inversions, duplications, and deletions), standard short-read sequencing cannot read the genome accurately. Long-read sequencing allows researchers to map the rearranged DNA blocks and gene copies, revealing the exact genetic changes causing the high CoA production.

Acknowledgement

I would like to express my deepest gratitude to my direct mentor, Warren Ho, for his invaluable guidance, patience, and continuous support throughout this project. I am equally grateful to my faculty supervisor, Patrick (Yizhi) Cai, for his insightful feedback, academic mentorship, and for providing the opportunity to conduct this research. Finally, I would like to thank everyone in the Cai Lab for creating a collaborative work environment and for their assistance at various stages of this study.

References

Barritt, S. A., DuBois-Coyne, S. E., & Dibble, C. C. (2024). Coenzyme A biosynthesis: Mechanisms of regulation, function and disease. *Nature Metabolism*, *6*(6), 1008–1023. https://doi.org/10.1038/s42255-024-01059-y

Boldogköi, Z. (2012). Transcriptional interference networks coordinate the expression of functionally related genes clustered in the same genomic loci. *Frontiers in Genetics*, *3*, 122. https://doi.org/10.3389/fgene.2012.00122

Ceccatelli Berti, C., Gihaz, S., Figuccia, S., Choi, J.-Y., Pal, A. C., Goffrini, P., & Ben Mamoun, C. (2022). Evidence for a Conserved Function of Eukaryotic Pantothenate Kinases in the Regulation of Mitochondrial Homeostasis and Oxidative Stress. *International Journal of Molecular Sciences*, *24*(1), 435. https://doi.org/10.3390/ijms24010435

Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwerzmann, J. W., Dai, J., Lindstrom, D. L., Boeke, A. C., Gottschling, D. E., Chandrasegaran, S., Bader, J. S., & Boeke, J. D. (2011). Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature*, *477*(7365), 471–476. https://doi.org/10.1038/nature10403

Gietz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, *2*(1), 31–34. https://doi.org/10.1038/nprot.2007.13

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 genes. *Science*, *274*(5287), 546, 563–567. https://doi.org/10.1126/science.274.5287.546

Hamza, A., Tammpere, E., Kofoed, M., Keong, C., Chiang, J., Giaever, G., Nislow, C., & Hieter, P. (2015). Complementation of Yeast Genes with Human Genes as an Experimental Platform for Functional Testing of Human Genetic Variants. *Genetics*, *201*(3), 1263–1274. https://doi.org/10.1534/genetics.115.181099

Iuso, A., Wiersma, M., Schüller, H.-J., Pode-Shakked, B., Marek-Yagel, D., Grigat, M., Schwarzmayr, T., Berutti, R., Alhaddad, B., Kanon, B., Grzeschik, N. A., Okun, J. G., Perles, Z., Salem, Y., Barel, O., Vardi, A., Rubinshtein, M., Tirosh, T., Dubnov-Raz, G., … Anikster, Y. (2018). Mutations in PPCS, Encoding Phosphopantothenoylcysteine Synthetase, Cause Autosomal-Recessive Dilated Cardiomyopathy. *The American Journal of Human Genetics*, *102*(6), 1018–1030. https://doi.org/10.1016/j.ajhg.2018.03.022

Kachroo, A. H., Laurent, J. M., Yellman, C. M., Meyer, A. G., Wilke, C. O., & Marcotte, E. M. (2015). Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science*, *348*(6237), 921–925. https://doi.org/10.1126/science.aaa0769

Koeppel, J., Ferreira, R., Vanderstichele, T., Riedmayr, L. M., Peets, E. M., Girling, G., Weller, J., Murat, P., Liberante, F. G., Ellis, T., Church, G. M., & Parts, L. (2025). Randomizing the human genome by engineering recombination between repeat elements. *Science*, *387*(6733), eado3979. https://doi.org/10.1126/science.ado3979

Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R., & Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genetics*, *39*(10), 1235–1244. https://doi.org/10.1038/ng2117

Leonardi, R., Zhang, Y.-M., Rock, C. O., & Jackowski, S. (2005). Coenzyme A: Back in action. *Progress in Lipid Research*, *44*(2), 125–153. https://doi.org/10.1016/j.plipres.2005.04.001

López, J., Bustos, D., Camilo, C., Arenas, N., Saa, P. A., & Agosin, E. (2020). Engineering Saccharomyces cerevisiae for the Overproduction of β-Ionone and Its Precursor β-Carotene. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.578793

Olzhausen, J., Grigat, M., Seifert, L., Ulbricht, T., & Schüller, H.-J. (2021). Increased biosynthesis of acetyl-CoA in the yeast Saccharomyces cerevisiae by overexpression of a deregulated pantothenate kinase gene and engineering of the coenzyme A biosynthetic pathway. *Applied Microbiology and Biotechnology*, *105*(19), 7321–7337. https://doi.org/10.1007/s00253-021-11523-4

Ruiz, A., González, A., Muñoz, I., Serrano, R., Abrie, J. A., Strauss, E., & Ariño, J. (2009). Moonlighting proteins Hal3 and Vhs3 form a heteromeric PPCDC with Ykl088w in yeast CoA biosynthesis. *Nature Chemical Biology*, *5*(12), 920–928. https://doi.org/10.1038/nchembio.243

Smith, J. J., Valentino, T. R., Ablicki, A. H., Banerjee, R., Colligan, A. R., Eckert, D. M., Desjardins, G. A., & Diehl, K. L. (2025). A genetically encoded fluorescent biosensor for visualization of acetyl-CoA in live cells. *Cell Chemical Biology*, *32*(2), 325-337.e10. https://doi.org/10.1016/j.chembiol.2025.01.002

Swidah, R., Auxillos, J., Liu, W., Jones, S., Chan, T.-F., Dai, J., & Cai, Y. (2020). SCRaMbLE-in: A Fast and Efficient Method to Diversify and Improve the Yields of Heterologous Pathways in Synthetic Yeast. *Methods in Molecular Biology*, *2205*, 305–327. https://doi.org/10.1007/978-1-0716-0908-8_17

Wang, J.-Y., Cao, Y.-Y., Chen, Y.-N., Wu, X.-L., He, B.-T., Zhu, S.-Y., Zhou, X., Wu, Y., Li, B.-Z., & Yuan, Y.-J. (2022). Artificial nondirectional site-specific recombination systems. *iScience*, *25*(1), 103716. https://doi.org/10.1016/j.isci.2021.103716

Zhang, Y., Wang, J., Wang, Z., Zhang, Y., Shi, S., Nielsen, J., & Liu, Z. (2019). A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in Saccharomyces cerevisiae. *Nature Communications*, *10*(1), 1053. https://doi.org/10.1038/s41467-019-09005-3

Zhang, Z., & Ren, Q. (n.d.). Why are essential genes essential? - The essentiality of Saccharomyces genes. *Microbial Cell*, *2*(8), 280–287. https://doi.org/10.15698/mic2015.08.218