Integration of metabolism and regulation reveals rapid adaptability to growth on non-native substrates.

Vikas D. Trivedi[1,2], Sean F. Sullivan[1], Debika Choudhury[1], Venkatesh Endalur Gopinarayanan[1], Taylor Hart[1], Nikhil U. Nair[1*]

[1] Department of Chemical & Biological Engineering, Tufts University, Medford, MA
[2] Current: Department of Structural Biology and Center for Data Driven Discovery, St. Jude Children's Research Hospital, Memphis, TN

* Corresponding author: nikhil.nair@tufts.edu; @nair_lab

ABSTRACT

Engineering synthetic heterotrophy (i.e., growth on non-native substrates) is key to efficient bio-based valorization of various renewable and waste substrates. Among these, engineering hemicellulosic pentose utilization has been well-explored in *Saccharomyces cerevisiae* (yeast) over several decades – yet the answers remain elusive about what makes their utilization inherently recalcitrant in yeast. Through implementation of a semi-synthetic regulon (REG), we find that harmonizing cellular and engineering objectives is key to obtaining highest growth rates and yields with minimal metabolic engineering effort showing that "extrinsic" factors – specifically, upstream genes that direct flux of pentoses into central carbon metabolism – are rate-limiting. We also reveal that yeast metabolism is innately highly adaptable to rapid growth on non-native substrates and that systems metabolic engineering (i.e., flux balancing, directed evolution, functional genomics, and network modeling) is largely unnecessary. We posit that the need for extensive engineering espoused by prior works is a consequence of unfortunate (albeit avoidable) antagonism between engineering and cellular objectives. We also found that deletion of endogenous genes to promote growth demonstrate inconsistent outcomes that are genetic-context- and condition-dependent. For the most part, these knockouts also lead to deleterious pleiotropic effects that decrease the robustness of strains against inhibitors commonly found in lignocellulosic hydrolysate. Thus, at best, perturbation of intrinsic factors (e.g., native metabolic, regulatory genes) results in incremental and inconsistent benefits. At worst, they are detrimental. Overall, this work provides insight into the limitations and pitfalls to realizing efficient synthetic heterotrophy using traditional/systems metabolic engineering approaches, demonstrates the innate adaptability of yeast for metabolism of non-native substrates, and provides an alternate, novel, holistic (and yet minimalistic) approach based on integrating non-native metabolic genes with a native regulon system.

KEYWORDS: synthetic biology, systems biology, metabolic engineering, sustainability, biomanufacturing

INTRODUCTION

Engineering metabolism for growth on non-native substrates (i.e., synthetic heterotrophy) has been an outstanding challenge for several decades in various microbial species (*1-4*). The traditional approach is to constitutively overexpress catabolic genes that input substrate to central carbon metabolism with the expectation that the native cellular systems can direct subsequent steps required for growth. In case initial designs do not lead to rapid growth, flux balancing (*5*), adaptive laboratory evolution (ALE) (*6, 7*), functional genomics (*8, 9*), and directed evolution (*6, 10*) approaches are used. More recently, systems-level analysis of regulatory structures in engineered strains have been used to identify dysregulated pathways and corrective interventions have provided avenues to improve engineering outcomes (*11-13*). Unifying to these approaches is that they are reactive interventions that try to rectify inefficiencies introduced due to strain engineering but not proactive in circumventing undesirable outcomes by harmonizing recombinant activity with native cellular objectives. Typical of this approach is engineering the yeast *Saccharomyces cerevisiae* for complete and rapid utilization of lignocellulosic pentoses (*14-17*). In this work, we engineer strains of yeast for rapid growth and pentose (xylose, arabinose) utilization and assess intrinsic and extrinsic constraints and pitfalls that control desirable phenotypic outcomes. Specifically, we answer the longstanding unresolved question: *what are the inherent/intrinsic limitations within this yeast that prevent it from metabolizing non-native pentose substrates at rates equivalent to native substrates?* Surprisingly, we find that this yeast is intrinsically highly adaptable to rapid growth on non-native substrates and that it can achieve maximum growth rates with minimal engineering – which is in direct contrast with the currently accepted paradigm.

To achieve this, we first modified our semi-synthetic regulon system (*18*) – based on synthetically activating the galactose (GAL) system by the non-native substrate – for efficient utilization of arabinose through the isomerase pathway (*19*). We demonstrate that while initial outcomes – like with xylose, previously – are superior to constitutively overexpressing the same genes, growth rates were modest ($\mu = 0.14 – 0.17\ h^{-1}$). To identify factors that constrain growth on either pentose, we performed pathway balancing, directed evolution, and systems biology-driven functional genomics. To our surprise, we found that when cells can coordinate global growth responses with substrate use through the semi-synthetic GAL regulon (REG), growth is largely extrinsically controlled – i.e., limited by the upstream/heterologous pathway that controls non-native substrate uptake and flux to central carbon metabolism. We also found through parallel investigations that genetic interventions identified through traditional/systems metabolic engineering that prune cellular metabolic and/or regulatory networks to improve strain performance often led to pleiotropic defects with mid-to-severe fitness costs. Thus, our findings suggest that an approach that repurposes native substrate utilization GAL regulon and synergizes it with an optimized upstream heterologous metabolic module reveals the hidden metabolic adaptability of yeast.

RESULTS

A semi-synthetic regulon design outperforms a constitutive overexpression strategy.

Previously, we demonstrated that coupling activation of galactose (GAL)-responsive regulon to catabolism and growth on non-native substrate, xylose, enabled faster growth and more complete utilization when compared to the constitutive overexpression of the same genes. This is because GAL regulon activation upregulates growth responsive genes and suppresses starvation responses, in contrast to the traditional constitutive overexpression approaches where the opposite holds true (**Figure 1A**). We wanted to assess whether the benefit was unique to xylose or if growth on other substrates, like arabinose, could also benefit from activation of the GAL regulon. Since the Gal3p$^{Syn4.1}$ was engineered to activate the GAL regulon in a xylose-dependent manner, we wondered if it could be activated by arabinose since the two sugars are structurally similar. We assessed activation of *GAL1p-EGFP* by wild-type Gal3p or engineered Gal3p$^{Syn4.1}$ in the presence of native Gal2p and/or engineered Gal2p$^{2.1}$ permease (*18*). We observed that a Gal3p$^{Syn4.1}$ and Gal2p$^{2.1}$ co-expressing strain showed activation on arabinose with low background and high dynamic range (**Figure 1B**). Interestingly, arabinose showed higher activation of *GAL1p-EGFP* than xylose, even though both Gal3p$^{Syn4.1}$ and Gal2p$^{2.1}$ were engineered for activity on xylose (**Figure 1C**). Given robust activation, we constructed a semi-integrant strain by integrating accessory genes including, sensor (*GAL3$^{Syn4.1}$*), transporter (*GAL2$^{2.1}$*), and transaldolase (*TAL1*) – expressed under GAL promoters – to generate a "REG" (regulon) strain background (**Figure 1D**). Similarly, as control, and to compare the with the traditional engineering approach, we integrated *GAL2$^{2.1}$* and *TAL1* under strong constitutive promoters to generate the "CONS" parental strain. We then assessed the growth of both strains transformed with *araBAD* or *XYLA*3-XKS1* arabinose and xylose, respectively. In ARA-REG and XYL-REG all genes were expressed under GAL-responsive promoters whereas in ARA-CONS and XYL-CONS all genes were under constitutive promoters (**Figure 1D**). On both carbon sources, strains that used the GAL regulon to coordinate substrate assimilation with global metabolism demonstrated higher growth rates (μ: ARA-REG = 0.14 $\pm$ 0.04 h^{-1}, XYL-REG = 0.17 $\pm$ 0.01 h^{-1}) compared to those that constitutively overexpressed the same genes (μ: ARA-CONS = 0.06 $\pm$ 0.01 h^{-1}, XYL-CONS = 0.11 $\pm$ 0.01 h^{-1}) in addition to higher final cell densities (OD_{600}: ARA-REG = 6.01 $\pm$ 1.62, ARA-CONS = 1.31 $\pm$ 0.04, XYL-REG = 10.8 $\pm$ 0.62, XYL-CONS = 5.3 $\pm$ 0.49) (**Figure 1E-F**). We next wanted to assess what factors prevented these REG strains from achieving maximum aerobic growth rate (μ = 0.22 – 0.29 h^{-1}) (*13, 20*). We explored two potential limitations – i) intrinsic, i.e., native metabolic or regulatory systems downstream of substrate uptake, and ii) extrinsic, i.e., upstream metabolic modules (primarily heterologous genes) that direct non-native substrates to central carbon metabolism.

Perturbation of intrinsic (downstream) factors result in modest and inconsistent improvements.

In traditional/systems metabolic engineering strategies, adaptive laboratory evolution (ALE) or rational data-driven approaches are often used to improve growth rates of initial strain designs, since growth rates and biomass yields are most often suboptimal. We decided to use a (data-driven) systems biology approach to identify "intrinsic" genetic targets to modify so that may lead to improved growth rates of

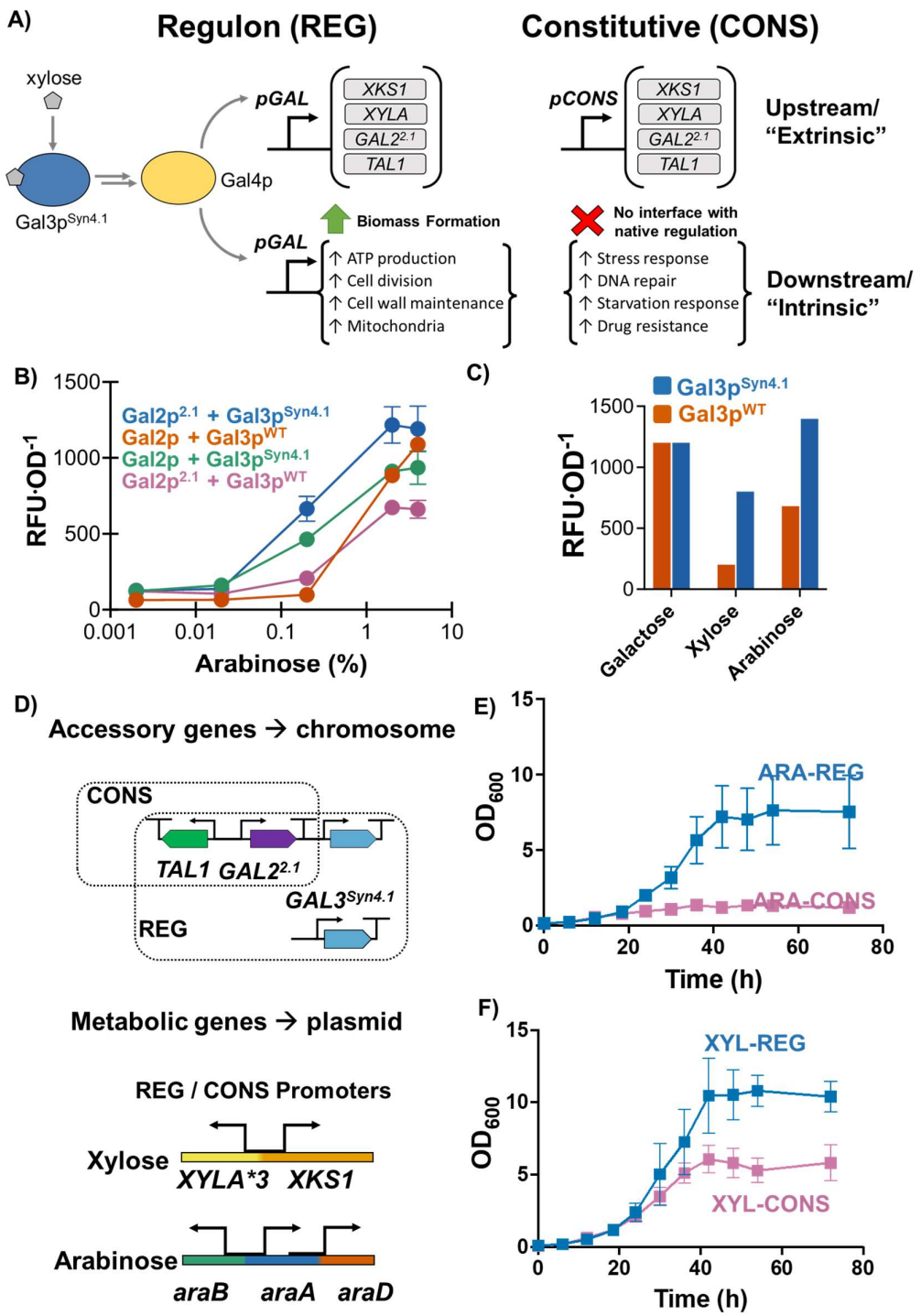


Figure 1. Implementation of a semi-synthetic regulon design for synthetic heterotrophy. A) The GAL regulon, synthetically activated by Gal3p$^{Syn4.1}$, regulates GAL-responsive genes as well as hundreds of downstream genes to promote growth. The traditional approach of constitutive overexpression cannot activate downstream genes, which leads to starvation response and slow growth (*18*). **A)** Dose-dependent *GAL1p-EGFP* activation by arabinose of strain expressing Gal3p^{WT} or Gal3p$^{Syn4.1}$ in and Gal2p^{WT} or Gal2p$^{2.1}$. **B)** Comparison of *GAL1p* activation by 2 % of native sugar, galactose, and non-native sugars – xylose and arabinose. **C)** Genetic structure of semi-integrant REG and CONS strain backgrounds. Growth curve of REG and CONS strains on **D)** arabinose and **E)** xylose when transformed with respective arabinose and xylose isomerase pathway genes, respectively. REG and CONS strains use GAL-responsive and constitutive promoters, respectively, for all genes indicated (**Table S1 and S2**).

our REG strains on xylose and arabinose (i.e., pentoses). By “intrinsic” factors, we mean native yeast genes that are far downstream of substrate uptake and catabolism (e.g., TCA cycle, DNA replication). For this, we identified differentially expressed genes between pentoses (non-native substrates) and galactose (a native substrate) – since growth rate on the latter has evolved to function harmoniously as part of the GAL regulon. Identifying the differences between two and closing the gap between them may aid in better synchronizing GAL regulation with pentose metabolism. We chose not to investigate previously identified genetic targets due to the vast difference in the transcriptional phenotype of REG strains relative to traditional CONS-type designs (*18*).

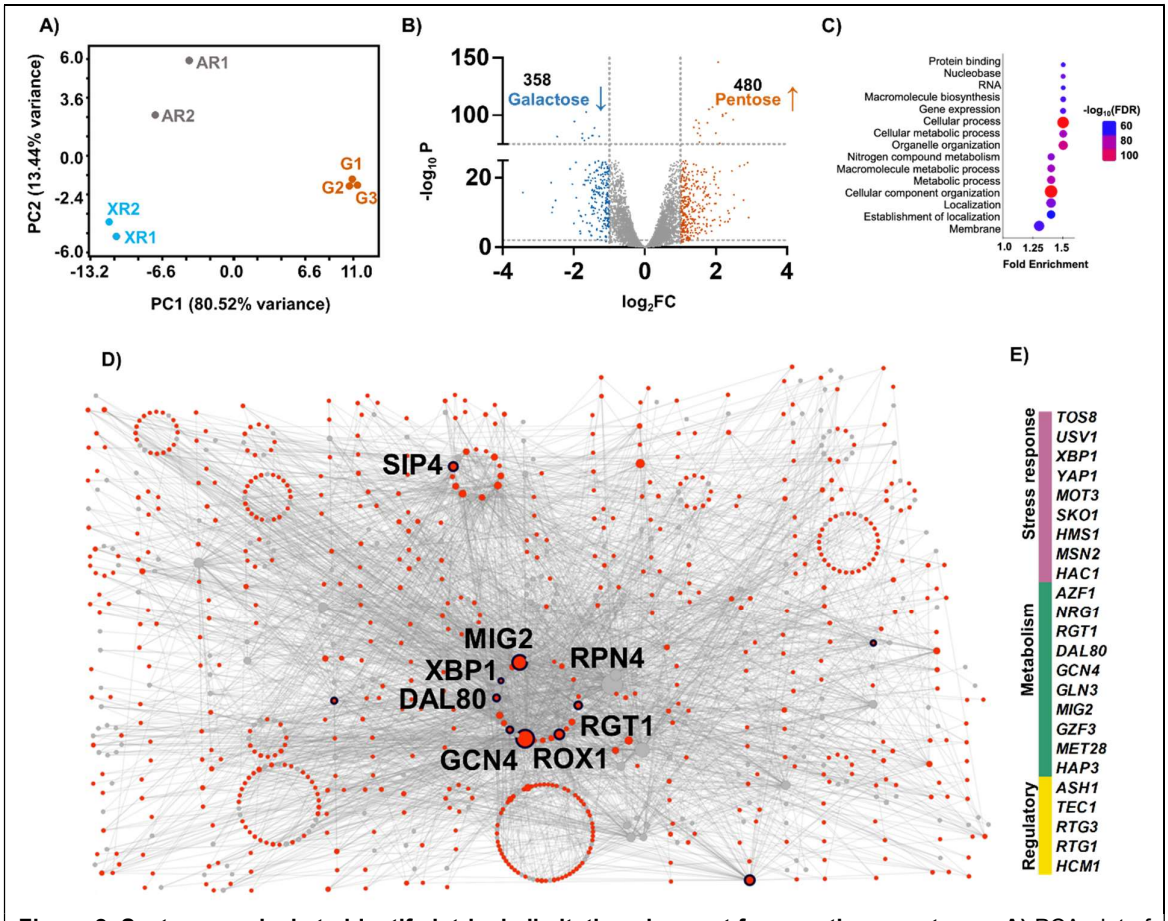


Figure 2. Systems analysis to identify intrinsic limitations in yeast for growth on pentoses. A) PCA plot of the three conditions AR#; ARA-REG grown on arabinose, XR#; XYL-REG grown on xylose, G#; WT grown on galactose. **B)** Differential expression volcano plot for transcriptome profile of galactose vs. xylose and arabinose combined. **C)** GO terms enriched among non-differentially expressed genes in the galactose vs. combined pentoses comparison are associated with growth. **D)** Network topology obtained by mapping results of differential gene expression analysis onto the EGRIN yeast gene regulatory network. **E)** 24 highly connected genes identified via differential gene expression and gene regulatory network analyses (with $p < 0.05$) as targets for deletion to improve growth on pentoses.

To identify targets, we performed RNA-seq on strains grown on arabinose (ARA-REG), xylose (XYL-REG), and galactose (WT) to characterize their respective expression profiles. As the GAL regulon has evolved to function harmoniously during growth on a native substrate, we hypothesized that identifying expression differences between galactose and the two non-native substrates and closing the gap

between them may aid in better synchronizing GAL regulation with pentose metabolism. Principal component analysis (PCA) of the transcriptomic profiles revealed clear separation of all three substrates, though the two pentoses clustered closer together than with the hexose, galactose, in PC1 (that represented 80.5% of the observed variance) (**Figure 2A**). A differential gene expression (DGE) analysis also revealed that a total of 865 genes were differentially regulated between ARA-REG and GAL-REG strains and 1455 genes between XYL-REG and GAL-REG strains (**Figure S1**) (p-value < 0.05 after Benjamini–Hochberg correction). We reasoned genes that are either directly regulated by GAL regulon (e.g., *GAL2, GAL4,* etc.) make a small fraction of the transcriptome and would not be differentially expressed in pentoses and hexoses. Rather than pursue the individual substrate comparisons, we chose to repeat the DGE analysis to compare pentose (XYL-REG and ARA-REG) vs. galactose utilization to identify "intrinsic" factors that differentiate growth on pentoses vs. a hexose. Upon doing so, we found 480 genes that were upregulated and 358 genes that were downregulated genes in pentoses vs. galactose (**Figure 2B**). Interestingly, many of the GO terms enriched among non-differentially expressed genes in the galactose vs. combined pentoses comparison are associated with growth, as expected from REG approaches (**Figure 2C**). Conversely, our previous comparison of CONS and REG approaches to growth on xylose and galactose showed upregulation of stress and starvation responses in CONS strains *(18)*.

Given the large number of differentially expressed genes (838) between the two conditions (pentoses vs. galactose), we sought to use gene regulatory networks (GRN) to identify core, highly-connected, regulatory elements that could serve as major contributors to the observed divergence in transcriptional profiles and therefore growth rates. We first identified three different yeast GRNs in the literature – EGRIN (*21*), YEASTRACT (*22*), and CLR (*23*) – constructed from diverse data types. Next, we mapped the differential expression data of each gene onto its corresponding node in each of the three networks. We then restricted the networks to just the significantly differentially expressed genes ($\log_2 > 1.5$, p-value < 0.05) and calculated the Betweenness Centrality value for each of the remaining nodes, collecting the identities of the top nodes. These highly-connected nodes (**Figure 2D**) broadly related to six gene ontology (GO) terms (**Figure S2**). Of these, 24 genes were chosen as targets for knock-out to test their effect on growth and phenotype (**Figure 2E**). We generated barcoded deletion (KO) libraries of these "intrinsic" factors in wild-type and REG strain backgrounds and performed growth enrichments on glucose, galactose, xylose, and arabinose (**Figure 3A**). Comparing population shifts using barcodes, we identified eight genes that displayed positive fitness on both xylose and arabinose but not on galactose – indicating a role in controlling growth primarily on pentoses. We then tested their growth individually on galactose, xylose, and arabinose (**Figure 3B-D**) and found that only Δ*GLN3* was either neutral or beneficial on both pentoses – all other deletions were detrimental for growth on xylose. Further, the growth rates of all knockout strains were still lower than that of the parental strain on galactose ($\mu = 0.24 \pm 0.03\ h^{-1}$ when *GAL1-7-10* are expressed from a plasmid, **Figure S5**), indicating additional limitations. Overall, these systems metabolic engineering approaches failed to identify genetic factors broadly beneficial to growth on pentoses. In contrast, the genome-wide expression changes effected by GAL regulon activation are beneficial for all three substrates.

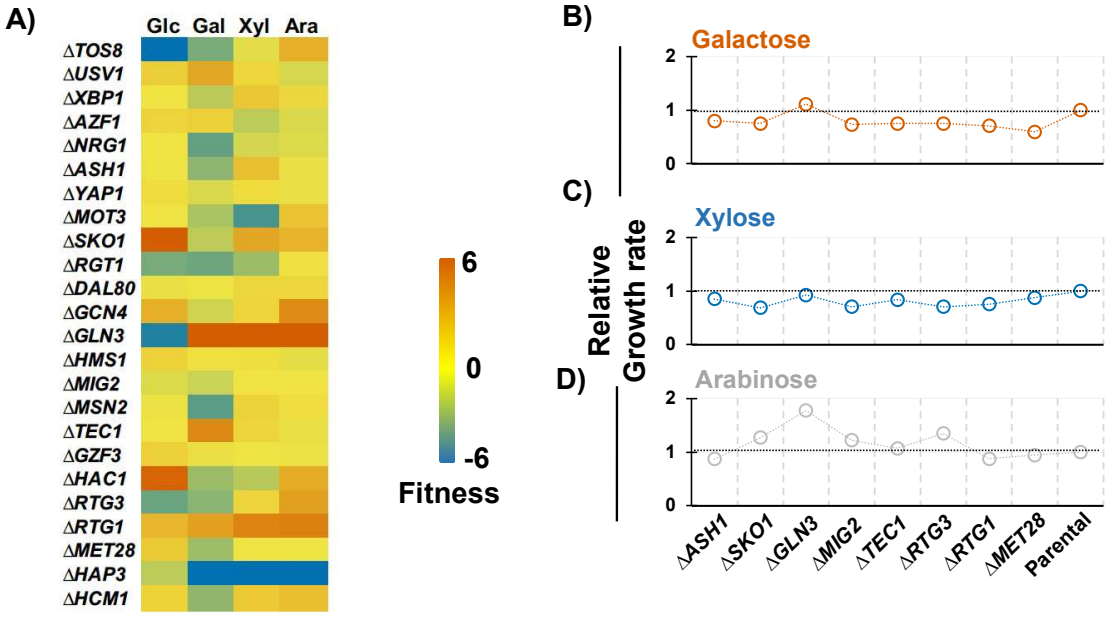


Figure 3. Effect of gene knockouts on growth phenotype. **A)** Fitness map of knock-out library of 24 genes identified from the mutual information network. The scale represents fitness ranging from blue (negative) to orange (positive). Growth rates of KOs with positive fitness on **B)** galactose, **C)** xylose (*GAL1p- XYLA*3 GAL10p -XKS1*), **D)** arabinose (*GAL1p-araB GAL10p-araA GAL7p-araD*).

Pentose metabolism is largely extrinsically (upstream) controlled with a regulon approach.

Our studies have so far focused on identifying "intrinsic" factors that may be controlling/bottlenecking growth on pentoses and given the inconsistent benefits, we wondered whether the limitations were, in fact, "extrinsic". The former implies that the native metabolic and/or regulatory capacity of this yeast is inherently limited for effective pentose metabolism and improving growth rate requires vast restructuring of associated (intrinsic) metabolic and/or regulatory networks. The latter implies that there are no inherent limitations in this yeast and that it is already poised for rapid growth on pentoses, but the observed low growth rates are due to suboptimal design of the upstream metabolic module that includes the heterologous (extrinsic) genes. To assess whether the "extrinsic limitation" paradigm has any merits, we needed to optimize the design of the upstream metabolic module responsible for substrate uptake and flux into central carbon metabolism (i.e., glycolysis).

First, we looked at the effect of plasmid copy number. For arabinose, there was no difference in growth rate when *araA-araB-araD* genes were expressed on high- or low-copy plasmids, whereas for xylose, we only observed growth when the *XYLA*3-XKS1* gene dose was high (**Figure S4**). In either case, changing plasmids backbone did not improve growth rate. Next, we hypothesized that balancing expression of these heterologous genes may be required to enhance growth rate. For arabinose, we created all six combinations of gene-promoter pairings and assessed their performance. Surprisingly, we found marked improvement in growth rate – from 0.14 ± 0.04 h^{-1} in the original design to 0.27 ± 0.03 h^{-1} in the best re-design (**Figure 4A**). To understand the cause of this behavioral change, we compared the relative expression levels of the *araA*, *araB,* and *araD* using quantitative reverse transcription PCR

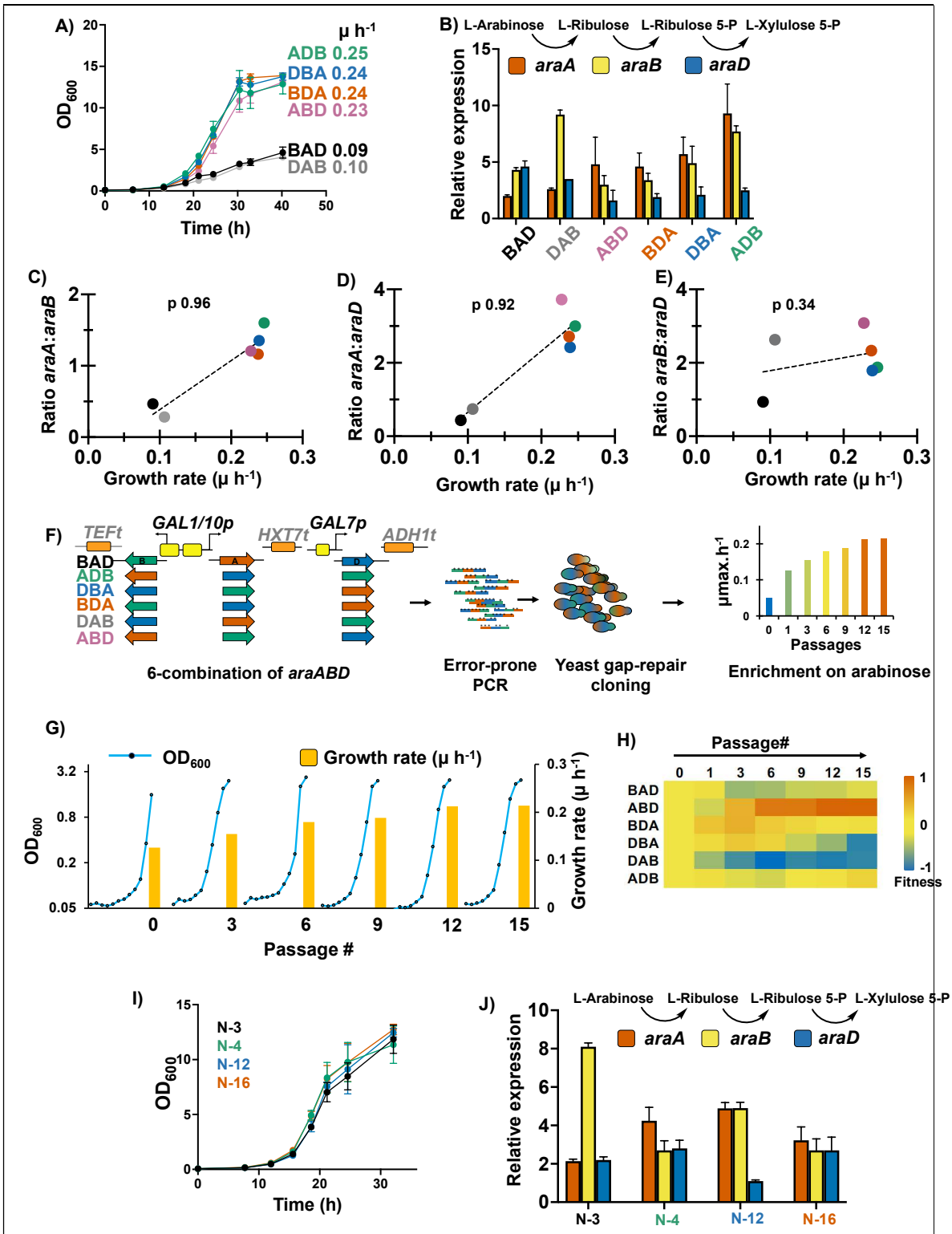
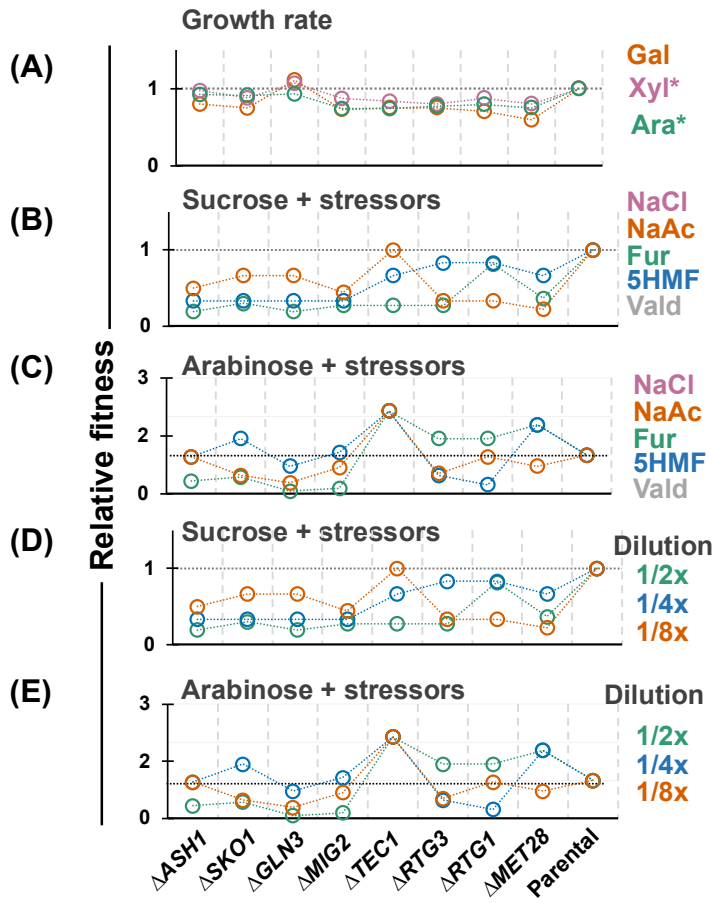


Figure 4. Optimizing extrinsic (upstream) factors to enhance growth of yeast on pentoses. A) Comparison of growth performance of different promoter-gene combination. **B)** Gene expression analysis of the six strains by qPCR. Pearson correlation coefficient between growth rate and gene expression levels of **C)** *araA:araB*, **D)** *araB:araD,* and **E)** *araA:araD*, respectively. **G)** Enrichment of mutagenic *araBAD* library in SC+Ara. **H)** Fitness heatmap of strain dynamics during enrichment on arabinose, reveled by sequencing barcodes. **I)** Growth curves and **J)** qPCR expression profile of *araBAD* genes encoded in plasmids isolated from top performing strains isolated after directed evolution.

228 (qPCR) (**Figure 4B**). We observed that in the poor performing combinations, expression of *araB*
229 (ribulokinase) was the highest, whereas, in high performing combinations, expression level followed
230 this pattern *araA* > *araB* > *araD*. In addition, we found a strong positive correlation between growth rate
231 and relative expression of *araA*:*araB* and *araA*:*araD*, respectively (**Figure 4C-E**). The success of this
232 approach encouraged us to attempt the same on xylose and found similar improvements – from 0.17 ±
233 0.01 h^{-1} to 0.24 ± 0.01 h^{-1} (**Figure S5**). These are comparable to the growth rate of yeast on galactose
234 when *GAL1-7-10* are expressed from a plasmid (0.24 ± 0.03 h^{-1}) (**Figure S3**).
235



236
237 **Figure 5**. **Effect of gene deletions on growth and stress-response phenotypes. A)** Growth rate of the optimized
238 plasmids in KO-strains. Fitness of KOs in presence of individual stressors when grown on **B)** sucrose and **C)**
239 arabinose or in the mixture of stressors on **D)** sucrose and **E)** arabinose. The relative fitness is expressed as growth
240 of deletion strains over the fitness of parental strain under the same conditions. NaCl = sodium chloride; NaAc =
241 sodium acetate; Fur = furfural; 5HMF = 5-hydroxymethylfurfural; Vald = veratraldehyde. 1/2x, 1/4x, 1/8x = serial
242 dilutions of the five stressors combined. Xyl* = *GAL10p-XYLA*3 GAL7p-XKS1*; Ara* = *GAL1p-araA GAL10p-araB*
243 *GAL7p-araD* in respective REG backgrounds.
244

245 Finally, we used directed evolution to improve the growth rate on arabinose further. We randomly
246 mutagenized the six arabinose pathway combinations, adding barcodes to track the lineages and
247 enriched this library size of 10^8 variants in minimal (SC+Ara) and complex (2YP+Ara) arabinose media.
248 Over the course of 15 subcultures, the growth rate over subcultures increased from 0.12 h^{-1} to 0.22 h^{-1}
249 and 0.18 h^{-1} to 0.26 h^{-1} in SC+Ara and 2YP+Ara, respectively (**Figure 4G, S6**). Using barcodes, we

250 tracked the performance of promoter-gene combinations. We observed that initially all the six plasmids
251 start at similar abundance, but the *araA-B-D* (under *GAL1-10-7p*, respectively) was the most abundant
252 at the end of enrichment in SC+Ara. We picked single colonies from each condition and calculated their
253 growth rates and identified 4 variants that showed the highest growth rates (0.35 ± 0.04 h^{-1}) from the
254 SC+Ara enriched culture (**Figure S7**). We sequenced the barcodes to identify the lineage and the whole
255 cassette to identify the mutations and found that three of the six initial designs were represented in the
256 four best variants (named N-3, N-6, N-12, and N-16) (**Table S3**). Re-transformation into parent
257 background strain indicated that the four variants attained the same maximum growth rate as that on
258 galactose (0.29 ± 0.01 h^{-1}), which is the highest reported growth rate on any pentose reported in the
259 literature for this yeast (**Figure 4I**). Since the mutations were distributed throughout the cassettes, we
260 quantified the expression levels of *araA*, *araB*, and *araD* in the four strains (**Figure 4J**) and found
261 patterns that differed significantly than parental strains (**Figure 4B**). We saw a strong positive
262 correlation between growth rate and relative expression of *araA*:*araB* and negative correlation between
263 *araB*:*araD*, indicating that the directed evolution campaign likely altered both activity and expression in
264 each strain differently (**Figure S8**). These results highlight a key insight about yeast: it's ability to utilize
265 pentoses is largely limited only by "extrinsic" factors (upstream pathway, especially heterologous
266 enzyme activity) and minimally by any "intrinsic" factor (i.e., native regulation or metabolic pathway).
267
268 **Engineering intrinsic (downstream) genes lead to pleiotropic fitness trade-offs.**
269 Next, we assessed whether the global regulatory elements identified through our network analysis
270 would benefit our best designs (*araA-B-D* under *GAL1-10-7p* and *XYLA*3-XKS1* under *GAL1-7p*).
271 Deleting each of the 8 genes in this strain did not result in any improvements over the parental strains
272 (**Figure 5A**). Here too, we could explore a combinatorial deletion of ALE strategy to enhance growth
273 rate. However, given that each strain with had already attained aerobic growth rate equivalent to the
274 maximum described for this yeast on glucose and galactose (0.22 h^{-1} – 0.29 h^{-1}) (*13, 20*) with high
275 biomass yields and short lag, we expect that improvements would be insignificant.
276
277 One consideration not yet accounted for is the suitability these strains for bioprocessing and their
278 resilience to growth inhibitors found in lignocellulosic hydrolysates. Given that all deletion targets are
279 highly-connected genes that control key cellular processes, we were concerned that dysregulation
280 major networks may lead to undesirable pleiotropic effects. Since resilience against stress is a complex
281 phenotype, often requiring concerted response from gene networks, we tested the fitness of all single
282 knockout (KO) strains on several stressors (sodium chloride, sodium acetate, furfural, 5-
283 hydroxymethylfurfural, and veratraldehyde) relevant to bioprocessing individually and in combination
284 (**Figure 5B-E**). Using parental strain as reference, we observed that performance of strains under stress
285 was highly context-dependent. For example, in sucrose medium, all deletions lost fitness in single or
286 mixed inhibitor cultures (**Figure 5C, E**). Conversely, on arabinose (with GAL regulon activated), certain
287 KO strains had improved tolerance to stressors (e.g., *ΔTEC1, ΔMET28*) (**Figure 5D, F**). *ΔGLN3* has
288 previously been shown to improve fitness under isobutanol stress (*24*); however, in our study, it was
289 less fit than the parental strain under all stress conditions. Interestingly, it did demonstrate improved

290 growth rate in a sub-optimal upstream metabolic design in arabinose (**Figure 3E**) but lost that benefit
291 in a more optimized design (**Figure 5C, E**). Collectively, these results highlight that deleting genes to
292 remodel expression profiles to enhance a single phenotype (e.g., growth rate on a non-native substrate)
293 can lead to some improvements, but they are often accompanied by negative pleiotropic effects that
294 make the strain less suitable for eventual bioprocessing applications where growth conditions are often
295 non-ideal.
296
297 **DISCUSSION**
298 Despite decades of effort engineering synthetic heterotrophy with *S. cerevisiae*, there is not formalized
299 understanding of what limits metabolic adaptability of this yeast for growth on pentoses. It is important
300 to differentiate between substrate utilization/uptake from growth since the former can be readily
301 achieved by diverting flux to unwanted or dead-end byproducts (e.g., Crabtree metabolites, organic
302 acids, pentitols) often with native substrate supplementation to support biomass generation. However,
303 valorization to high-value products requires efficient catabolism to central metabolic products whose
304 concentrations and fluxes are regulated and associated with growth. Initial studies in literature focused
305 on upstream ("extrinsic") elements (e.g., heterologous gene expression/activity, transporter
306 engineering, etc. (*25-29*)) whereas recent focus has been more "intrinsic", focusing on functional
307 genomics, adaptively laboratory evolution (ALE), and often network/systems analysis, to identify
308 downstream limitations (*5-13*). Successes have been abundantly reported with improvements in growth
309 along with a series of deletion and overexpression targets (*PHO13, ALD6, ASK10, YPR1, SNF6, RGT1,*
310 *CAT8, MSN4, GPD1, CCH1, ADH6, BUD21, ALP1, ISC1, RPL20B, COX4, ISU1, SSK2, YLR042c,*
311 *CYC8, PHD1, TEC1, ARR1,* etc. (*13, 30-36*)). Interestingly, many targets are unique to individual
312 studies, further highlighting the context dependence of targets on initial strain designs. Typifying the
313 traditional approach is a recent report where all the aforementioned extensive systems metabolic
314 engineering approaches were used to develop a strain of yeast with high growth aerobic growth rate
315 (0.26 h^{-1}) and short lag phase (9-15 h) on xylose (*13*). However, it is not clear if the insights were
316 translatable to engineering growth on any other substrate.
317
318 Through our work, we posit that requirement for extensive engineering to identify intrinsic limitations is
319 only required when the initial strain design is sub-optimal. Indeed, constitutive overexpression of
320 upstream metabolic genes results in significant stress (*18*) and low growth rates that must be
321 compensated for through systems approaches focused on preventing cellular detection of stress rather
322 than activating growth supporting systems. We demonstrated that with an optimal upstream module,
323 strains can attain superior growth profiles (fast specific growth rate, $\mu \geq 0.24\ h^{-1}$; short lag phase, <10
324 h; high final biomass, OD_{600} >10) if a growth-associated regulon – the GAL regulon – is activated. This
325 builds on our prior observation that synthetic activation of the GAL regulon potentiates cells for rapid
326 growth on non-native substrates by upregulating growth responsive genes and suppressing starvation
327 responses (*18*). Importantly, using this approach our final strain designs had minimal modifications –
328 we deleted only one gene (other than the Leloir pathway genes) – *GRE3* – to minimize oxidation of
329 substrate pentoses to pentitols. Many genes that have previously been identified as important

330 inactivation targets to improve growth on pentoses are intact. We also overexpressed only a minimal
331 upstream metabolic module (*TAL1* and *GAL2*$^{2.1}$) along with the specific pentose metabolic genes
332 (*araBAD* or *XYLA*3-XKS1*). Our work strongly argues that growth on pentoses is largely extrinsically
333 controlled and there are no major intrinsic regulatory or metabolic limitations in this yeast. This insight
334 presents a paradigm shift in engineering synthetic heterotrophy.
335
336 An additional advantage of this approach is the preservation of native regulatory systems that are
337 otherwise dysregulated in the traditional engineering approach. We found that while deletion of
338 endogenous genes could improve growth in strains with sub-optimal upstream modules under certain
339 conditions, they are associated with pleiotropic defects and are less robust in the presence of common
340 lignocellulosic inhibitors. This is not surprising since most genetic interventions aim to suppress cellular
341 responses to stress rather than promote growth. Thus, strains engineered for synthetic heterotrophy
342 through extensive gene inactivation are less robust for eventual bioprocess applications. This highlights
343 a major drawback of the traditional systems metabolic engineering approaches like ALE. In contrast,
344 our regulon-based approach is minimalistic and holistic – our strains maintain native regulatory systems
345 and even exploited them toward the engineering goal (i.e., the GAL regulon). This approach significantly
346 simplifies and expedites the design-build-test cycle for synthetic heterotrophy and demonstrates the
347 intrinsic adaptability of yeast toward growth on non-native substrates. We expect that this insight and
348 approach will expand the utility of this yeast for valorizing current and emerging waste and/abundant
349 substrates.
350
351 **DATA AVAILABILITY**
352 RNA-seq sequencing data has been submitted to NCBI SRA and is available under accession
353 PRJNA837644. R script for mutual information network is available at https://github.com/nair-lab/yeast-
354 MINet.
355
356 **ACKNOWLEDGEMENTS**
357 The authors would like to thank current and former Nair lab members, Dr. Todd C. Chappell, and Dr.
358 Karishma Mohan for helpful discussions. This work was supported by NIH grant #DP2HD91798, NSF
359 grant #1935354, and Tufts Launchpad | Accelerator to N.U.N.
360
361 **AUTHOR CONTRIBUTIONS**
362 N.U.N., V.D.T., and D.C. conceived and designed the research project. V.D.T., S.F.S., D.C., and N.U.N.
363 co-wrote the manuscript. V.D.T., S.F.S., V.E.G., and T.H. performed the experiments. V.D.T., S.F.S.,
364 and N.U.N. analyzed the data. All the authors have reviewed the manuscript and approved it for
365 submission.
366
367 **COMPETING INTERESTS**
368 None.
369

370 **MATERIALS AND METHODS**

371

372 **Strains and plasmids.**

373 The list of plasmids and strains used are listed in Supplementary **Table S1 and S2**. Strain W303-1a
374 (*MATa leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15*) and plasmids pIS374, pIS376, pIS385
375 were obtained from Euroscarf (Germany). The plasmids were constructed in the present using NEB-
376 HiFi DNA assembly master mix from NEB (Beverly, MA).

377

378 **Growth studies.**

379 Overnight inoculums were grown in the required dropout SC medium with sucrose (2 %). The culture
380 was washed twice in the growth medium and resuspended at an initial OD_{600} of 0.1 with appropriate
381 sugar (2 %) in 250 mL shake flasks containing 20 mL of media. OD600 measurement was checked at
382 frequent time intervals (3-6 h) on SpectraMax M3 spectrophotometer (Molecular Devices). Growth rate
383 was determined by plotting the values in GraphPad Prism following non-linear regression and using
384 exponential growth equation, $Y = Y_0 \exp(kX)$.

385

386 **Directed evolution.**

387 Error prone PCR libraries of arabinose cassettes (pVDT14, pVDT38-42) were generated as described
388 earlier (*18*). Six-barcoded primers (**Table S4**) were used to track the proportion of the lineage during
389 the enrichment. The amplicons from error prone PCR were assembled into pRS413 plasmids using
390 yeast gap-repair cloning. To attain high library size, the linear fragments were transformed into yeast
391 by electroporation as described previously (*37*). We attained library size of ~10^8 CFUs which was then
392 subjected to enrichment on arabinose in 2YP as well SC media. The cell pellets from each passage
393 were frozen for quantifying barcodes using amplicon sequencing (Genewiz, Cambridge, MA). At the
394 end of enrichment, the library pool was plated on 2YP+Ara and SC+Ara and 18 colonies were randomly
395 picked for growth rate determination. The plasmids from best variants were isolated from strain,
396 transformed into *E. coli*. The corresponding plasmid isolated from *E. coli* was sequenced and re-
397 transformed into REG-strain for growth rate determination.

398

399 **Genomics integrations.**

400 Accessory cassettes (pVDT29, 30, 35, **Table S1**) were integrated into VEG16 (W303-1a, Δ*GRE3*,
401 Δ*GAL1,* Δ*GAL3,* Δ*GAL7,* Δ*GAL10*) (*18*) strain via disintegrator plasmid(*38*). Counterselection was
402 performed on SC medium with 1 g/L 5-FOA to remove the *URA3* marker and the locus was amplified,
403 and sequence confirmed.

404

405 **Deletion library.**

406 To generate the deletion library, the cassettes for the required targets were amplified from the genomic
407 DNA of the appropriate strains from the KO collection (*39*). Since the cassettes contained *KANMX*
408 marker, the library was selected on YP+G418 (400 µg/mL). The pool was stored as a stock and used
409 for studying the fitness by transforming it with plasmids for galactose, xylose, and arabinose utilizing
410 cassettes.

411

412 **qPCR expression analysis.**

413 Total yeast RNA isolation was performed on OD ~ 1 of yeast cells collected from mid-log phase growth
414 using the GeneJet RNA Purification Kit (Thermofisher Scientific Catalog #: K0731) according to the
415 instructions for isolating yeast RNA. Two samples were collected for each growth condition. The
416 concentration of total RNA isolated was obtained by measuring 10-fold dilutions from each sample on
417 a spectrophotometer. Using these concentrations, 2.5 µg of total RNA was subjected to the 'routine'
418 DNase treatment protocol using the Invitrogen Turbo DNA-free Kit (Thermofisher Scientific Catalog #:
419 AM1907). Next, 20 % of the final reaction (10-50 µL, targeting 500 ng RNA) was used to synthesize
420 cDNA for the sample using the Invitrogen SuperScript IV First-Strand Synthesis System (Thermofisher
421 Scientific Catalog #: 18091050). The random hexamers supplied by the kit were used to prime the
422 reverse transcriptase enzyme. cDNA samples were diluted 100-fold (2 µL into 198 µL dH_2O), of which
423 2 µL was used as the template for a 15 µL qPCR reaction using the Applied Biosystems PowerUp
424 SYBR Green Master Mix (Thermofisher Scientific Catalog #: A25741). Each sample was tested for the
425 expression level of five genes: the three arabinose catabolic genes (*araA, araB*, and *araD*) as well as
426 two housekeeping genes (*TFC1* and *UBC6*) using the primers listed in **Table S5**. Reactions were run
427 on an Applied Biosystems Quantstudio 5 Real-Time PCR Instrument. The relative expression level of
428 each catabolic genes was calculated by subtracting the geometric mean of the cycle thresholds (Ct) of
429 the two housekeeping genes from the Ct of that catabolic gene, followed by conversion of the log-base
430 two Ct into a non-log value that can be compared across genes.

431

432 **RNA-seq.**

433 Transcriptomics of strains WT, XYL-REG, ARA-REG were performed on the mid-log phase cultures
434 grown on their respective carbon source (galactose, xylose, or arabinose). Cells pellets were washed
435 twice in water and stored at −80 °C and outsourced to Genewiz Inc. for RNA extraction and sequencing.
436 RNA-seq was performed on Illumina HiSeq. Raw FASTQ files were processed for differential
437 expression analysis using GeneiousPrime. The possible adapter sequences and low-quality short-read
438 (less than 50 bp) trimming were performed using BB-Trim package. The reads were aligned to the
439 reference genome W303 obtained from Saccharomyces Genome Database
440 (http://www.yeastgenome.org) using Bow-Tie2 package. The edgeR package was used to normalize
441 the gene count based on library size and was converted to cpm (counts per million) using. DESEQ2
442 package was sued for differential gene expression analysis. Genes with p-values < 0.05 and fold
443 change of ≥ 2 were considered as differentially expressed.

444

445 **Amplicon sequencing analysis.**

446 The total DNA was extracted and quantified using a spectrophotometer and approximately 100 ng was
447 used as template for PCRs. Unique, barcoded primers flanked by Illumina sequencing adaptors were
448 used to generate the amplicons after 15-20 cycles of PCR for sequencing (**Table S6**). The barcoded
449 samples were pooled and sent for amplicon sequencing (2×250 bp, Genewiz, New Jersey, USA). For
450 each pooled sample, we received approximately 100,000 sequencing reads. These data were

processed according to a previously described bioinformatic workflow using Geneious Prime® 2020.2.4 (*40*). Briefly, the reads were paired and merged using the BBMerge package and filtered for poor-quality reads using the BBDuk package. The reads were mapped using BowTie2 and gene expression was calculated and differential expression was determined using DESeq2 (*41*).

Network analysis.

The *S. cerevisiae* genes that we consider as the 'gold standard' list of transcription factors is the unique set of genes identified in the file supplemental data **File S1**. Our analysis utilizes three yeast genetic regulatory networks (EGRIN, YEASTRACT, and CLR) (*42*) in the context of integrating regulatory and metabolic networks to enable more accurate prediction of yeast phenotypes in different growth conditions (*43*). The Environment and Gene Regulatory Influence Network (EGRIN) consists of 92 regulators and 2,588 interactions and was constructed using the cMonkey and Inferelator computational tools trained on expression data from 2,929 microarray experiments to identify gene clusters and potential regulatory genes. The YEASTRACT gene regulatory network was extracted directly from the YEASTRACT database2 and consists of 177 regulators and 31,075 regulatory associations based on both 'direct evidence' – chromatin immunoprecipitation (ChIP), ChIP-on-chip, electrophoretic mobility shift assay, or examination of the effect of TF binding site mutations on target gene expression – as well as 'indirect evidence' provided by gene expression changes in response to deletion, mutation, or overexpression of a given TF3.

The context likelihood of relatedness (CLR) algorithm quantifies the amount of mutual information shared between all possible gene pairs in each set of gene expression data. The regulatory network is then inferred by taking the distribution of calculated mutual information for all gene pairs and filtering out gene pairs whose mutual information content is below a chosen threshold with the remaining pairs considered to be connections in the network4. The R programming language was used for importing and manipulating data. R scripts can be found in https://github.com/nair-lab/yeast-MINet.

Stressor studies.

We determined fitness of the strains in presence of various stressors encountered during biomanufacturing using plate assay. Briefly, we tested the growth of by performing spot dilution on 2YP supplemented with 2 % sucrose or arabinose and various concentration of the stressors (**Table S7**). The plates after incubation were imaged and the colonies were counted to determine the CFUs. The CFUs in the presence of the stressor was divided by the CFUs in absence of the stressor to determine the fitness of that strain. This value was normalized to the fitness of the parental strain and was expressed as the relative fitness.

487 **REFERENCES**

488 1. J. Sun, W. Sun, G. Zhang, B. Lv, C. Li, High efficient production of plant flavonoids by microbial
489 cell factories: Challenges and opportunities. *Metabolic Engineering*, (2022).
490 2. M. A. Franden *et al.*, Engineering Pseudomonas putida KT2440 for efficient ethylene glycol
491 utilization. *Metabolic engineering* **48**, 197-207 (2018).
492 3. S. Pontrelli *et al.*, Escherichia coli as a host for metabolic engineering. *Metabolic engineering*
493 **50**, 16-46 (2018).
494 4. T. Sun, Y. Yu, K. Wang, R. Ledesma-Amaro, X.-J. Ji, Engineering Yarrowia lipolytica to produce
495 fuels and chemicals from xylose: A review. *Bioresource Technology* **337**, 125484 (2021).
496 5. T. J. Hanly, M. A. Henson, Dynamic flux balance modeling of microbial co-cultures for efficient
497 batch fermentation of glucose and xylose mixtures. *Biotechnology and bioengineering* **108**,
498 376-385 (2011).
499 6. S.-M. Lee, T. Jellison, H. S. Alper, Directed evolution of xylose isomerase for improved xylose
500 catabolism and fermentation in the yeast Saccharomyces cerevisiae. *Applied and*
501 *Environmental Microbiology* **78**, 5708-5716 (2012).
502 7. L. Diao *et al.*, Construction of fast xylose-fermenting yeast based on industrial ethanol-
503 producing diploid Saccharomyces cerevisiaeby rational design and adaptive evolution. *BMC*
504 *biotechnology* **13**, 1-9 (2013).
505 8. D. J. Wohlbach *et al.*, Comparative genomics of xylose-fermenting fungi for enhanced biofuel
506 production. *Proceedings of the National Academy of Sciences* **108**, 13212-13217 (2011).
507 9. A. Matsushika, H. Inoue, T. Kodaki, S. Sawayama, Ethanol production from xylose in
508 engineered Saccharomyces cerevisiae strains: current state and perspectives. *Applied*
509 *microbiology and biotechnology* **84**, 37-53 (2009).
510 10. M. Lee, H. J. Rozeboom, E. Keuning, P. de Waal, D. B. Janssen, Structure-based directed
511 evolution improves S. cerevisiae growth on xylose by influencing in vivo enzyme performance.
512 *Biotechnology for Biofuels* **13**, 1-18 (2020).
513 11. Z. Dai, J. Nielsen, Advancing metabolic engineering through systems biology of industrial
514 microorganisms. *Current opinion in biotechnology* **36**, 8-15 (2015).
515 12. K. V. Presnell, H. S. Alper, Systems metabolic engineering meets machine learning: A new era
516 for data-driven metabolic engineering. *Biotechnology journal* **14**, 1800416 (2019).
517 13. X. Li *et al.*, Metabolic network remodelling enhances yeast's fitness on xylose using aerobic
518 glycolysis. *Nature Catalysis* **4**, 783-796 (2021).
519 14. L. Sun, Y. S. Jin, Xylose assimilation for the efficient production of biofuels and chemicals by
520 engineered Saccharomyces cerevisiae. *Biotechnology Journal* **16**, 2000142 (2021).
521 15. X. Wang, J. Yang, S. Yang, Y. Jiang, Unraveling the genetic basis of fast l-arabinose
522 consumption on top of recombinant xylose-fermenting Saccharomyces cerevisiae.
523 *Biotechnology and Bioengineering* **116**, 283-293 (2019).
524 16. R. Garcia Sanchez *et al.*, Improved xylose and arabinose utilization by an industrial
525 recombinant Saccharomyces cerevisiae strain using evolutionary engineering. *Biotechnology*
526 *for biofuels* **3**, 1-11 (2010).
527 17. H. W. Wisselink, M. J. Toirkens, Q. Wu, J. T. Pronk, A. J. van Maris, Novel evolutionary
528 engineering approach for accelerated utilization of glucose, xylose, and arabinose mixtures by
529 engineered Saccharomyces cerevisiae strains. *Applied and environmental microbiology* **75**,
530 907-914 (2009).
531 18. V. Endalur Gopinarayanan, N. U. Nair, A semi-synthetic regulon enables rapid growth of yeast
532 on xylose. *Nature communications* **9**, 1-12 (2018).
533 19. H. W. Wisselink *et al.*, Engineering of Saccharomyces cerevisiae for efficient anaerobic
534 alcoholic fermentation of L-arabinose. *Applied and Environmental Microbiology* **73**, 4881-4891
535 (2007).
536 20. Z. Dai, M. Huang, Y. Chen, V. Siewers, J. Nielsen, Global rewiring of cellular metabolism
537 renders Saccharomyces cerevisiae Crabtree negative. *Nature communications* **9**, 1-8 (2018).
538 21. A. N. Brooks *et al.*, A system-level model for the microbial regulatory genome. *Molecular*
539 *systems biology* **10**, 740 (2014).
540 22. M. C. Teixeira *et al.*, The YEASTRACT database: a tool for the analysis of transcription
541 regulatory associations in Saccharomyces cerevisiae. *Nucleic acids research* **34**, D446-D451
542 (2006).
543 23. J. J. Faith *et al.*, Large-scale mapping and validation of Escherichia coli transcriptional
544 regulation from a compendium of expression profiles. *PLoS biology* **5**, e8 (2007).
545 24. K. Kuroda *et al.*, Critical roles of the pentose phosphate pathway and GLN3 in isobutanol-
546 specific tolerance in yeast. *Cell Systems* **9**, 534-547. e535 (2019).

25. E. Young, S.-M. Lee, H. Alper, Optimizing pentose utilization in yeast: the need for novel tools and approaches. *Biotechnology for biofuels* **3**, 1-12 (2010).
26. E. M. Young, A. D. Comer, H. Huang, H. S. Alper, A molecular transporter engineering approach to improving xylose catabolism in Saccharomyces cerevisiae. *Metabolic engineering* **14**, 401-411 (2012).
27. J. G. Nijland, A. J. Driessen, Engineering of pentose transport in Saccharomyces cerevisiae for biotechnological applications. *Frontiers in Bioengineering and Biotechnology*, 464 (2020).
28. B. Wiedemann, E. Boles, Codon-optimized bacterial genes improve L-arabinose fermentation in recombinant Saccharomyces cerevisiae. *Applied and environmental microbiology* **74**, 2043-2050 (2008).
29. V. Endalur Gopinarayanan, N. U. Nair, Pentose metabolism in Saccharomyces cerevisiae: the need to engineer global regulatory systems. *Biotechnology journal* **14**, 1800364 (2019).
30. S. R. Kim *et al.*, Rational and evolutionary engineering approaches uncover a small set of genetic changes efficient for rapid xylose fermentation in Saccharomyces cerevisiae. *PloS one* **8**, e57048 (2013).
31. J. Usher *et al.*, Chemical and Synthetic Genetic Array Analysis Identifies Genes that Suppress Xylose Utilization. (2011).
32. Y. Shen *et al.*, An efficient xylose-fermenting recombinant Saccharomyces cerevisiae strain obtained through adaptive evolution and its global transcription profile. *Applied Microbiology and Biotechnology* **96**, 1079-1091 (2012).
33. L. V. Dos Santos *et al.*, Unraveling the genetic basis of xylose consumption in engineered Saccharomyces cerevisiae strains. *Scientific Reports* **6**, 1-14 (2016).
34. N. S. Parachin, O. Bengtsson, B. Hahn-Hägerdal, M. F. Gorwa-Grauslund, The deletion of YLR042c improves ethanolic xylose fermentation by recombinant Saccharomyces cerevisiae. *Yeast* **27**, 741-751 (2010).
35. J. Zha, M. Shen, M. Hu, H. Song, Y. Yuan, Enhanced expression of genes involved in initial xylose metabolism and the oxidative pentose phosphate pathway in the improved xylose-utilizing Saccharomyces cerevisiae through evolutionary engineering. *Journal of Industrial Microbiology and Biotechnology* **41**, 27-39 (2014).
36. C. F. Wahlbom, R. R. Cordero Otero, W. H. van Zyl, B. r. Hahn-Hägerdal, L. J. Jönsson, Molecular analysis of a Saccharomyces cerevisiae mutant with improved ability to utilize xylose shows enhanced expression of proteins involved in transport, initial xylose metabolism, and the pentose phosphate pathway. *Applied and environmental microbiology* **69**, 740-746 (2003).
37. J. R. Thompson, E. Register, J. Curotto, M. Kurtz, R. Kelly, An improved protocol for the preparation of yeast cells for transformation by electroporation. *Yeast* **14**, 565-571 (1998).
38. I. Sadowski, T. C. Su, J. Parent, Disintegrator vectors for single-copy yeast chromosomal integration. *Yeast* **24**, 447-455 (2007).
39. G. Giaever *et al.*, Functional profiling of the Saccharomyces cerevisiae genome. *nature* **418**, 387-391 (2002).
40. V. D. Trivedi *et al.*, In-Depth Sequence–Function Characterization Reveals Multiple Pathways to Enhance Enzymatic Activity. *ACS Catalysis* **12**, 2381-2396 (2022).
41. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 1-21 (2014).
42. Z. Wang *et al.*, Combining inferred regulatory and reconstructed metabolic networks enhances phenotype prediction in yeast. *PLoS computational biology* **13**, e1005489 (2017).
43. M. J. Herrgård, B.-S. Lee, V. Portnoy, B. Ø. Palsson, Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome research* **16**, 627-635 (2006).