

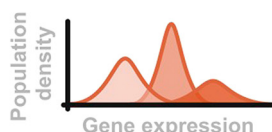
**Research Article**

# Promoter selection reduces phenotypic heterogeneity and improves bioproduction

**Classical  
Metabolic Engineering**

↓

Use of  
macro-heterogeneous promoters



↓

Strain engineered with  
macro-heterogeneous promoters  
produces subpopulations



↓

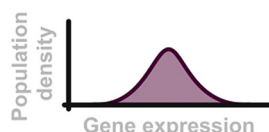
Poor bioproduction



**Flow cytometry + mixture modelling  
scRNA seq + regression analysis**

↓

Identification of  
homogeneous promoters



↓

Strain engineered with  
homogeneous promoters  
produces a single population



↓

Superior single-cell and  
bulk level bioproduction



Promoter macro-heterogeneity creates distinct subpopulations that limit bioproduction. Combining flow cytometry, scRNA-seq-based modelling, and single-cell Raman spectroscopy in *Saccharomyces cerevisiae*, we demonstrate that replacing macro-heterogeneous promoters with homogeneous alternatives improves yield up to threefold in a metabolic pathway, highlighting promoter heterogeneity as an important consideration for strain design.

Davina Patel, Kiyan Shabestary, Jiabao Xu (徐嘉宝), Cinzia Klemm, Wei E. Huang (黄巍), Rodrigo Ledesma-Amaro

[r.ledesma-amaro@imperial.ac.uk](mailto:r.ledesma-amaro@imperial.ac.uk)  
(R. Ledesma-Amaro).

## Highlights

Promoters in the yeast toolkit are macro-heterogeneous and form distinct subpopulations of high and low producer cells.

We developed a regression model to predict promoter macro-heterogeneity from scRNA-seq data which could be validated by flow cytometry.

Replacing macro-heterogeneous promoters with homogeneous promoters in a multistep pathway increased both single-cell production and bulk end-product yield by as much as threefold.

Our results suggest that macro-heterogeneous promoters are detrimental to bioproduction and should be avoided when building metabolic pathways.

Trends in Biotechnology, Month 2026,  
Vol. xx, No. xx  
<https://doi.org/10.1016/j.tibtech.2026.08.002>

## Research Article

## Promoter selection reduces phenotypic heterogeneity and improves bioproduction

Davina Patel <sup>1</sup>, Kiyan Shabestary <sup>1,5</sup>, Jiabao Xu  (徐嘉宝)<sup>2,6</sup>, Cinzia Klemm <sup>1</sup>, Wei E. Huang (黄巍)<sup>2</sup>, and Rodrigo Ledesma-Amaro <sup>1,3,4,\*</sup>

Cell-to-cell heterogeneity leads to the formation of subpopulations within an isogenic population, resulting in the coexistence of high and low producer cells. However, the impact of this phenotypic heterogeneity on bioproduction has been overlooked. Using flow cytometry to analyse commonly used yeast promoters, we found that many display macro-heterogeneity, generating multiple discrete subpopulations. Additionally, by regressing single-cell protein expression on single-cell RNA sequencing data, we explored the use of predictive models as a guide for selecting promoters with more homogeneous expression at desired expression levels. As a proof of concept, we engineered an improved lycopene pathway by substituting homogeneous promoters for macro-heterogeneous ones and implemented single-cell Raman spectroscopy to demonstrate up to a threefold increase in single-cell and bulk lycopene production. Overall, our results show that promoter heterogeneity can have a large impact on microbial bioprocesses and that the selection of homogeneous promoters can achieve more efficient production in a metabolically engineered pathway.

## Introduction

In strain engineering, genetic circuits and metabolic pathways are grafted into a host cell through the expression of heterologous proteins. In accordance with engineering principles, such multipart assemblies necessitate standardised parts whose performances have been independently characterised prior to assembly [1]. For biological systems, characterisation is critical due to inherent stochasticity, in which part performances can drastically differ among genetically identical cells and across conditions. In particular, differential gene and protein expression analyses have revealed pronounced cell-to-cell metabolic heterogeneity in both yeast and bacterial systems [2–7]. In addition to part unpredictability, metabolic heterogeneity can also adversely affect large-scale production in bioreactors where the emergence of low-producing subpopulations can outcompete high-producing strains [8,9]. Cell-to-cell variability may not be completely obvious and can be masked by a large spread in expression levels among cells [8,9]. This form of variation, termed **micro-heterogeneity** (see [Glossary](#)), is characterised by a large coefficient of variation (CV) in which expression values are continuously distributed and deviate widely from the population mean [8,9]. However, metabolic heterogeneity can also be present more obviously in the form of **macro-heterogeneity**. In this case, expression distributions are multimodal, with multiple peaks corresponding to high- and low-expressing cell populations [8,9].

One potential source of this variability is promoter heterogeneity, which can be influenced by both intrinsic and extrinsic factors. Sources of extrinsic noise include differences in cell morphology,

## Technology readiness

Promoter-driven phenotypic heterogeneity produces mixed populations of cells and reduces the efficiency of microbial cell factories. In this study, we identified macro-heterogeneity in commonly used YTK promoters, generated a model to predict promoter heterogeneity, and demonstrated that homogeneous promoters in a metabolically engineered pathway can increase the yield of the end product up to threefold. The promoter characterisation and regression model presented here form the basis of a promoter selection framework that is at technology readiness level 4 or 5, as we could demonstrate proof of concept by increasing yield in shake flask fermentations at laboratory scale. In future, subsequent validation in bioreactors is required to confirm productivity improvements associated with homogeneous promoters and to quantify the impact of promoter heterogeneity under scaled industrial fermentation conditions. Nevertheless, as sustainable microbial fermentations continue to gain importance, promoter homogenisation provides a practical strategy to improve the performance of engineered metabolic pathways. Compatible with existing metabolic engineering workflows, this approach can be readily applied across a range of microbial production platforms.

<sup>1</sup>Department of Bioengineering and Imperial College Centre for Synthetic Biology, Imperial College London, London SW7 2AZ, UK

<sup>2</sup>Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK

<sup>3</sup>Bezos Centre for Sustainable Protein, Imperial College London, London SW7 2AZ, UK

<sup>4</sup>UKRI Engineering Biology Mission Hub on Microbial Food, Imperial College London, London SW7 2AZ, UK

cell cycle stage, media composition, and genetic background, whereas sources of intrinsic noise can be traced back to the stochasticity of transcription [10,11]. In yeast, increased intrinsic promoter heterogeneity has been linked to TATA boxes within the promoter region [12,13], as well as transcriptional noise and promoter activation frequency [10]. Additionally, extrinsic factors such as media composition or cell cycle stage can influence promoter expression levels [14–19].

Given the key role of promoters as regulators of heterologous gene expression, research efforts have largely focused on methods to improve mean promoter expression beyond that of native promoters. Previously, promoter sequence [13,20–22], transcription factor binding [20], RNA polymerase binding [23], and nucleosome affinity [24,25] have been used to inform the rational design of strong synthetic promoters. While these studies have yielded valuable insights into promoter architecture and have highlighted ways to reduce promoter heterogeneity, they have largely overlooked the implications of this heterogeneity in downstream applications within more complex heterologous systems, such as those in metabolic pathways.

Advancements in **single-cell RNA sequencing (scRNA-seq)** have also significantly deepened our understanding of heterogeneity within **isogenic populations**, revealing how processes such as cell cycle regulation and carbon source utilisation contribute to phenotypic diversity—often as strategies to boost population fitness [5,26]. These studies provide valuable insights into transcriptional variability and predominantly emphasise technological developments in scRNA-seq or the effects of environmental stressors on expression [27–30]. However, as core regulators of gene expression, promoters themselves can exhibit heterogeneity within clonal populations [10–15]. Given the central role of promoter usage in metabolic engineering and bioproduction, where fine-tuned control of gene expression is essential, further studies to understand promoter-level variations in this context are required. Recent findings in bacterial expression systems point towards promoter heterogeneity and the emergence of subpopulations [31,32]; however, less is known about the effects of promoter heterogeneity on subpopulation formation in yeast and its knock-on effects on downstream metabolic engineering and bioproduction.

In this work, we studied promoter heterogeneity in the model organism *Saccharomyces cerevisiae*. As a case study, we selected some of the most widely used promoters in metabolic engineering and synthetic biology, specifically those from the **Yeast Toolkit (YTK)** [33,34]. The toolkit includes 19 native yeast promoters (pYTK) covering a wide range of expression levels, with the strongest constitutive promoter, pTDH3, showing more than a 50-fold increase in red fluorescence protein (RFP) fluorescence relative to pREV1, the weakest promoter in the toolkit [33]. We characterised each of the 19 constitutive promoters of the toolkit and identified macro-heterogeneity, wherein superfolder green fluorescent protein (sfGFP) expression was clustered into two to three subpopulations with differing sfGFP intensities. Following characterisation of promoter heterogeneity, we built a promoter selection guide in which promoter heterogeneity and strength could be assessed using single-cell transcript distributions derived from scRNA-seq datasets. We further showcased the utility of this guide by selecting homogeneous promoters to express the **lycopene** biosynthesis pathway, demonstrating that **promoter homogenisation** is beneficial for metabolic engineering. Taking advantage of the latest advances in **single-cell Raman spectroscopy** [35–37], we demonstrated a 2.9-fold increase in single-cell lycopene production following the selection of homogeneous promoters, along with increased bulk lycopene production. Together, our findings show that widely used promoters can form distinct subpopulations of high- and low-expressing cells, which can negatively affect the productivity of synthetic metabolic pathways. Consequently, promoter heterogeneity should be considered a key design principle for future microbial cell factories.

<sup>5</sup>Present address: Calico Life Sciences LLC, South San Francisco, CA 94080, USA

<sup>6</sup>Present address: Division of Biomedical Engineering, James Watt School of Engineering, University of Glasgow, Glasgow G12 8LT, UK

\*Correspondence:  
[r.ledesma-amaro@imperial.ac.uk](mailto:r.ledesma-amaro@imperial.ac.uk)  
(R. Ledesma-Amaro).

## Results

### Widely used promoters lead to macro-heterogeneous protein production

Promoters exhibit varying strengths of expression and noise, which may result in subpopulations of cells with distinct gene expression levels. To further investigate promoter heterogeneity at the single-cell level, 19 promoters from the YTK system [33] were used to control the expression of sfGFP in the *S. cerevisiae* laboratory strain BY4741 (Figure 1A; Tables S7 and S8 in the Supplemental information online). Golden Gate assembly was used to construct each pYTK–sfGFP transcriptional unit. Due to noise that could arise from differences in plasmid copy numbers, all constructs were integrated into the genome at the URA3 locus. Additionally, auxotrophy was cured using the pHLM minichromosome series, a centromeric plasmid carrying HIS3, LEU2 and MET17 [38]. Strains were grown in four common media (YPD10, YPD, YNB, and YEPG) (see ‘STAR★Methods’ section; Table S13 and Figures S15 and 16 in the Supplemental information online) and sampled at exponential and stationary phases to investigate single-cell heterogeneity at different growth stages using flow cytometry (Figure 1A). The distribution of sfGFP expression was analysed using **Hartigan’s dip test**, a statistical tool that uses *P*-values to assess multimodality, that is, the presence of multiple modes that may reflect distinct subpopulations of cells (Table S1 in the Supplemental information online). Following Hartigan’s dip test, the flexmix R package<sup>†</sup> as the creator/manufacture is Leisch et al. [39] was used to determine the number of subpopulations present, based on finite mixture modelling (Table S2 in the Supplemental information online).

Consequently, we identified a cluster of promoters that exhibited consistent macro-heterogeneity across various media conditions and growth phases (Figure 1B; Figures S1 and S2 in the Supplemental information online), which could not be identified using total sfGFP intensity measurements in a plate reader (Figure 1C). Notably, the promoters pHHF2, pHHF1, pHTB2, and pRNR1 demonstrated significant heterogeneity ( $P < 0.05$ , Hartigan’s dip test) in their expression profiles (Figure 1B,C; Table S1). Interestingly, we observed dynamic changes in the expression levels of several promoters depending on the growth phase and specific media conditions. For instance, the promoters pTDH3 and pRET2 exhibited significantly different levels of macro-heterogeneity in expression in YPD10 medium during the transition from the exponential to the stationary phase ( $P < 0.05$ , Hartigan’s dip test) (Figure 1B; Table S1). Remarkably, these two promoters showed consistent and uniform expression across all other media conditions during both growth phases, shedding light on context-dependent regulatory patterns underlying the activation of these promoters. In contrast, ten out of the 19 promoters tested did not exhibit any macro-heterogeneity across all growth phases and conditions tested. These were pPGK1, pTEF1, pTEF2, pALD6, pPAB1, pSAC6, pRNR2, pPOP6, pPSP2, and pREV1 ( $P < 0.05$ , Hartigan’s dip test) (Figure 1B; Table S1). Notably, there was a strong growth reduction in the pTDH3 strain when cultivated in YEPG medium (Figure 1B). We propose that this could be attributed to the growth burden caused by the strength of the promoter driving the expression of sfGFP; across all media conditions, the pTDH3 strain showed an extended time to mid-log of 30 min to 1 h compared with the other 18 strains tested (Figure S3 in the Supplemental information online). Furthermore, it reached a lower final OD<sub>600</sub> (Figure S4 in the Supplemental information online), suggesting that the high expression load imposed by pTDH3 might be too demanding for growth in YEPG, a medium devoid of glucose.

### The proportion of cells in each subpopulation reflects differences in cell cycle stage

After identifying macro-heterogeneous promoters, we next sought insights into the underlying causes of heterogeneity in promoters showing multimodal sfGFP expression. By categorising promoters into homogeneous, macro-heterogeneous, or mixed expression across all sample conditions, we compared the functions of the genes associated with each promoter based on

## Glossary

**FlowRACS:** high-throughput microfluidic system for Raman-based cell sorting. Cells enter a microfluidic chip and are trapped under a 532-nm laser to detect Raman fingerprints of individual cells, enabling cell sorting and the quantification of specific metabolites at the single-cell level.

**Gaussian process:** nonparametric regression method that fits functions directly to data without assuming a fixed mathematical form for the relationship between variables.

**Hartigan’s dip test:** tests whether a distribution has one peak or several by measuring how far the data deviate from the best single-peaked fit. A small *P* value (below a chosen threshold, typically 0.05) provides evidence against unimodality, so the population is likely multimodal.

**Isogetic populations:** subsets of cells that share an essentially identical genotype but differ in phenotype, reflecting nongenetic sources of variation.

**Lycopene:** carotenoid that serves as a reporter metabolite for quantifying bioproduction in an engineered metabolic pathway. In *S. cerevisiae*, lycopene is synthesised through heterologous expression of the CrtE, CrtB, and CrtI enzymes and accumulates as a red-coloured end product that can be quantified at single-cell resolution.

**Macro-heterogeneity:** form of expression variability characterised by a multimodal distribution in which peaks correspond to discrete subpopulations of cells occupying distinct expression states, as opposed to continuous variation within a single population.

**Micro-heterogeneity:** continuous variability in expression within a single population of cells, forming a unimodal distribution around the population mean rather than forming discrete subpopulations.

**Promoter homogenisation:** metabolic engineering strategy in which homogeneous promoters are selected to drive unimodal gene expression across a cell population, preventing the emergence of distinct subpopulations.

**Single-cell Raman spectroscopy:** analytical method used to determine the chemical composition of individual cells. A laser is focused onto a cell, and light is scattered by molecular bonds, producing a characteristic ‘Raman

their definitions in the *Saccharomyces* Genome Database (Table S3 in the Supplemental information online). Following gene lookup, DNA replication processes, such as chromatin assembly and DNA synthesis, were observed in macro-heterogeneous and mixed expression promoter groups (Table S3). While this could suggest a potential link to cell division, it is worth noting that larger datasets would be required to validate such a correlation. Furthermore, one major contributing factor to extrinsic noise and heterogeneity is the cell cycle stage. Past microarray data have shown that HHF1, HHF2, HTB2, and RNR1 transcript levels can vary greatly across the cell cycle [18]. Additionally, the proportion of cells in different stages of the cycle can influence gene expression intensity at the population level, especially when transcription factors respond to cyclin oscillations, causing periodic gene expression [18,19]. Therefore, the heterogeneity of these promoters may be linked to the presence of cells in various cell cycle stages.

In order to understand the impact of cell cycle stage on promoter heterogeneity, we sampled a subset of strains containing either a homogeneous or macro-heterogeneous promoter, in YPD, and treated them with  $\alpha$ -factor (AF)—a peptide which arrests cells in the G<sub>1</sub> stage of the cell cycle. Cells (Table S8) were grown for 6 h then either treated with AF for 2.5 h or left untreated to continue into the exponential phase. Samples were taken for flow cytometry and gated according to sfGFP levels to distinguish between low (Pop1) and high (Pop2) expressing populations (Figure 2A). Following AF treatment, cells expressing sfGFP under the control of homogeneous promoters (pPGK1, pALD6, and pREV1), all continued to homogeneously express sfGFP (Figure 2A). A slight increase in autofluorescence was found in those promoters as well as in the sfGFP-negative control, which could be explained by the change in cell morphology as AF induces shmooing, a specific phenotype in yeast, making cells larger and distorted to facilitate fusion while mating with a partner [40]. However, in cells expressing macro-heterogeneous promoters (pHHF1, pHHF2, pHTB2, and pRNR1), subpopulations of cells still arose, as illustrated by the presence of multiple peaks (Figure 2A). To quantify the impact of cell cycle arrest on subpopulation size, we gated the low- and high-expressing populations—Pop1 and Pop2, respectively (Figure 2A). Then, the proportion of cells in Pop1 and Pop2 were recorded before and after AF treatment (Figure 2B). For AF-treated cells, the proportion of cells in Pop1 increases by approximately 10% compared with untreated exponential phase (exp) cells ( $P < 0.001$ , Welch's  $t$ -test). The significant variation in subpopulation ratios after G<sub>1</sub> arrest suggests that cell cycle could be one of the factors playing a role in macro-heterogeneity.

To further assess whether promoter heterogeneity is derived from cell cycle state rather than solely from population-level arrest, we examined sfGFP expression at the single-cell level while explicitly assigning individual cells to specific cell cycle phases. Cells expressing sfGFP from either homogeneous or macro-heterogeneous promoters were stained with Hoechst to visualise DNA content and imaged by fluorescence microscopy. Using a combination of Hoechst signal distribution and cell morphology (budding state), individual cells were classified as being in G<sub>1</sub> or S/G<sub>2</sub>/M phase (Figure S5A in the Supplemental information online).

Cells expressing sfGFP from the homogeneous promoter pTEF1 displayed comparable sfGFP intensity across G<sub>1</sub> and S/G<sub>2</sub>/M populations, indicating that expression from this promoter is largely insensitive to cell cycle progression (Figure S5B). In contrast, strains expressing sfGFP from macro-heterogeneous promoters (pHHF1, pHHF2, pHTB2, and pRNR1) exhibited pronounced cell cycle-dependent differences in sfGFP abundance. For each of these promoters, sfGFP intensity was significantly altered between cells in G<sub>1</sub> and those in S/G<sub>2</sub>/M ( $P < 0.001$ ; Table S9 in the Supplemental information online), consistent with the emergence of subpopulations observed in bulk flow cytometry measurements (Figures S1 and S2).

'fingerprint' that reflects molecular vibrations. This fingerprint can then be used to identify and quantify intracellular molecules.

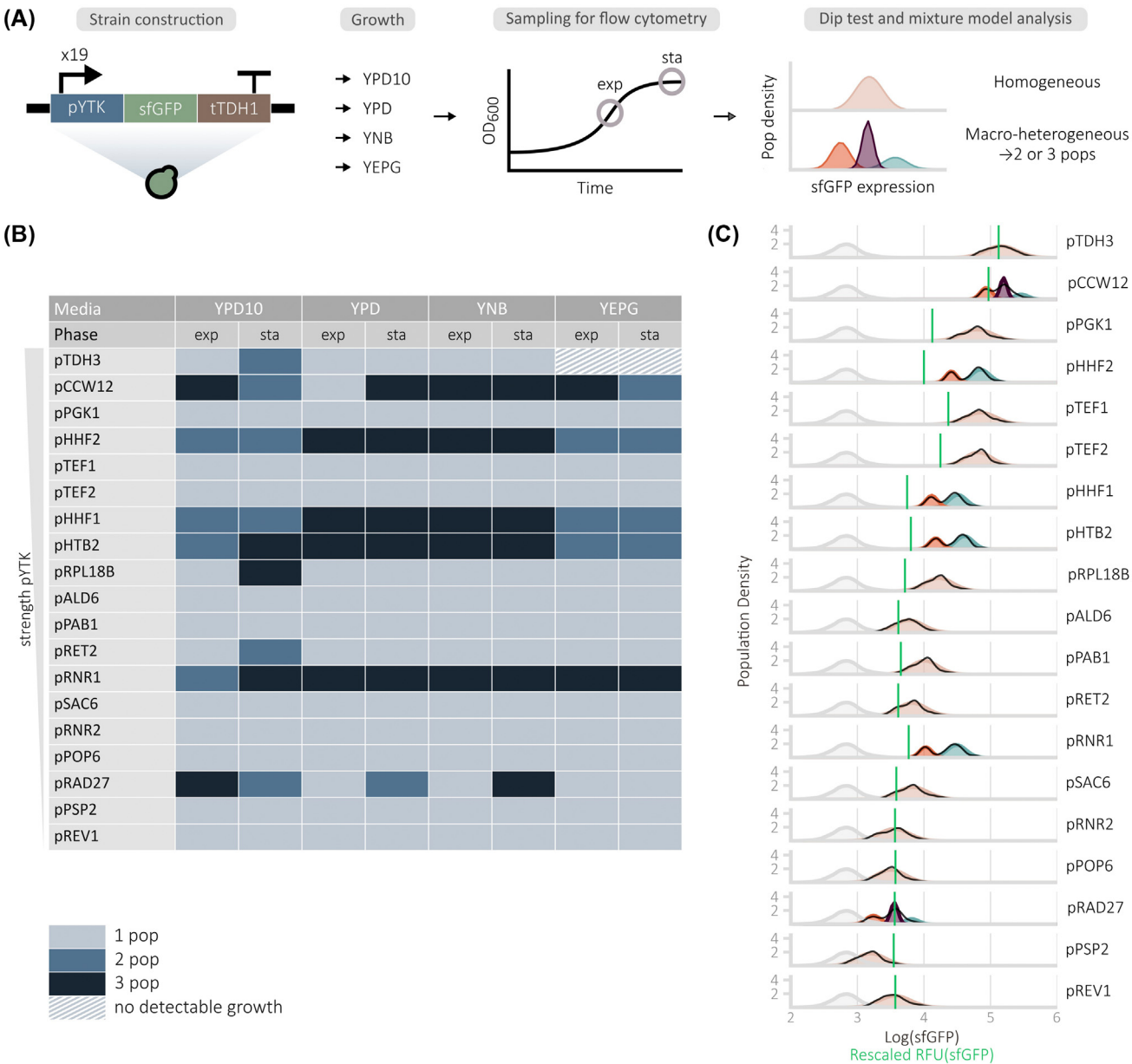
#### **Single-cell RNA sequencing**

**(scRNA-seq):** sequencing method that profiles the transcriptome of individual cells, measuring gene expression at single-cell resolution rather than as an average across a population, enabling detection of cell-to-cell heterogeneity (micro- or macro-scale).

#### **Uniform manifold approximation and projection (UMAP):**

general nonlinear dimensionality reduction technique that embeds high-dimensional data into a lower-dimensional space while preserving local structure and is commonly used in scRNA-seq to visualise cell-to-cell heterogeneity and identify clusters of transcriptionally similar cells.

**Yeast Toolkit (YTK):** standardised modular cloning system for building genetic constructs in *S. cerevisiae*. It relies on Golden Gate assembly to combine interchangeable DNA parts (e.g., promoters, coding sequences, and terminators) into expression cassettes, which can then be further assembled into multigene constructs. It was created by Lee and colleagues [7].



**Figure 1. Strain construction and characterisation of promoter macro-heterogeneity.** (A) Experimental framework. We generated 19 strains, each expressing sfGFP under the control of one constitutive promoter in the Yeast Toolkit (pYTK-sfGFP). All constructs were integrated into the genome at the URA3 locus. Strains were grown in four media conditions: YPD10, YPD, YNB, and YEPG, and samples were taken at exponential (exp) and stationary (sta) phases for flow cytometry. A dip test was performed on the resulting sfGFP distributions, and promoters were classified as macro-heterogeneous when  $P < 0.05$ . Mixture model analysis was then performed on macro-heterogeneous populations to determine whether two or three subpopulations were present. (B) Promoter macro-heterogeneity during exponential (exp) and stationary (sta) phases in four media conditions. Hartigan's dip test was used to identify macro-heterogeneous promoters ( $P < 0.05$ ; Table S1). These promoters were then analysed using a mixture model, and the number of subpopulations (two or three) was determined by selecting the model with the lowest BIC (Table S2). Promoters are ranked by reported strength, from highest to lowest [33]. Striped: no detectable growth; light blue: one population; medium blue: two subpopulations; dark blue: three subpopulations. (C) Overlay of two sfGFP measurement methods during exponential phase in YPD10: flow cytometry and global mean sfGFP expression (Figures S1 and S2). Global mean sfGFP expression was measured using a plate reader (green line) and scaled to align with flow cytometry data (see 'STAR★Methods' section). Black line: raw data; coloured fill: subpopulations determined via mixture modelling; grey fill: negative control. BIC: Bayesian

(Figure legend continued at the bottom of the next page.)

Together with the AF arrest experiments, these observations suggest that the cell cycle stage contributes substantially to macro-heterogeneous expression by selectively enriching cells with distinct transcriptional output. However, as subpopulations persist even under G<sub>1</sub> arrest and expression variability remains within individual cell cycle phases, additional regulatory mechanisms, such as transcriptional bursting, chromatin context, or fluctuating transcription factor availability, are likely to act in combination with cell cycle regulation to generate the observed heterogeneity.

#### scRNA-seq data can be a predictor of protein-level strength and heterogeneity

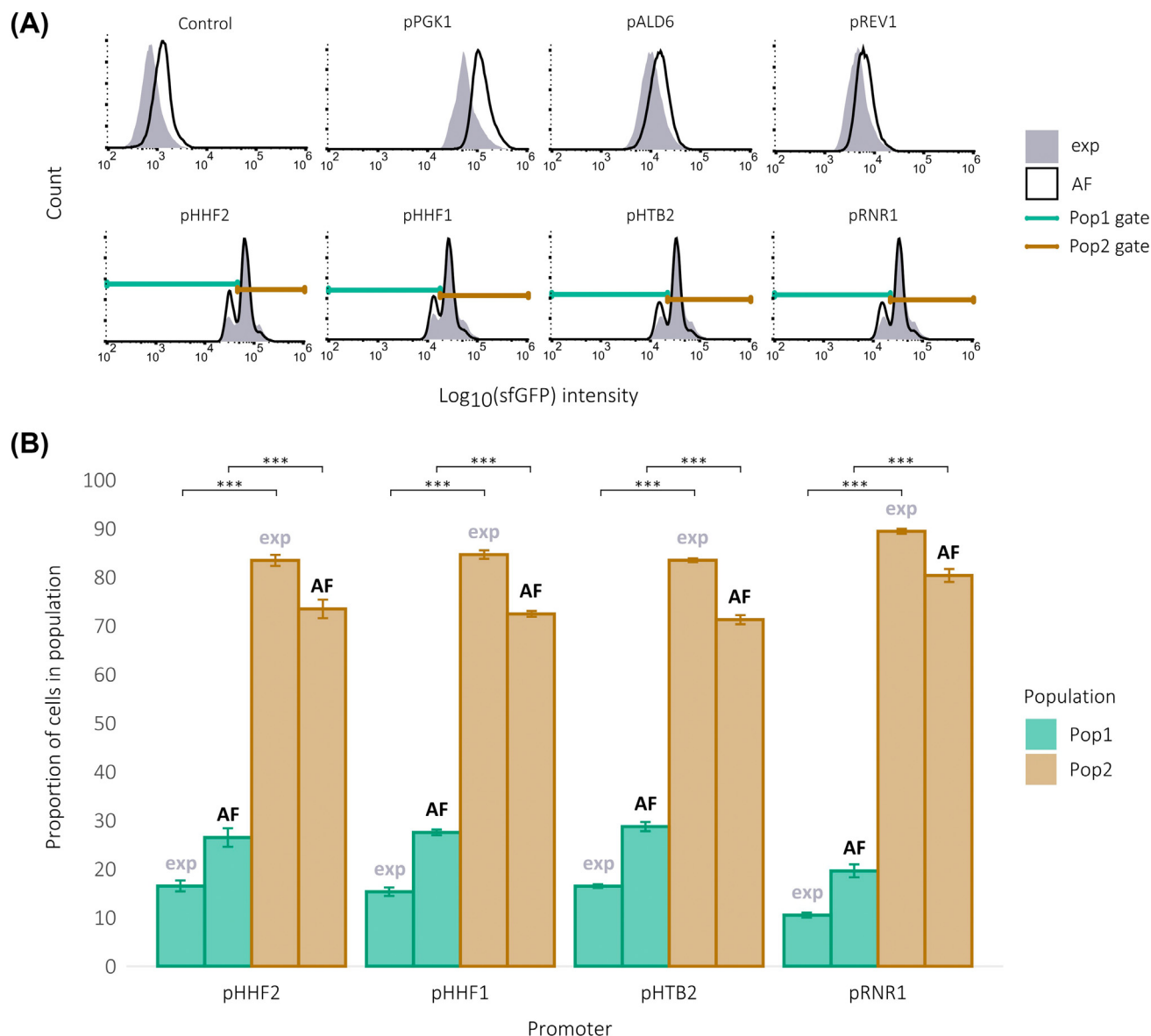
We next investigated whether promoter heterogeneity (both micro and macro) and strength in sfGFP expression could be predicted *a priori* based on scRNA-seq. scRNA-seq data have been used to characterise the emergence of subpopulations [5,41,42], but not to investigate promoter characteristics across diverse growth conditions. We decided to use scRNA-seq data from Jackson and colleagues [41], which include 38,000 cells across 11 environmental conditions, grown in rich (YPD) and minimal media (MMD). Leveraging this scRNA-seq dataset, we aimed to compare gene expression strength and noise with those from protein expression levels previously obtained by flow cytometry, for each of the 19 promoters of the YTK in minimal and rich media. For macro-heterogeneity, we applied Hartigan's dip test to both data types and found that it was not fully successful at identifying promoters that were macro-heterogeneous (Figure S6 and Table S4 in the Supplemental information online), suggesting that promoter heterogeneity cannot sufficiently inform on protein heterogeneity at the single-gene level and that a more holistic approach is required. Micro-heterogeneity was quantified as the CV, defined as the standard deviation (SD) divided by the mean expression across cells at both the protein (flow cytometry) and transcript (scRNA-seq) levels. Promoter strength was defined as the corresponding mean expression values. Our results indicate a moderate to good correlation between scRNA-seq data and sfGFP expression data for both promoter strength (Spearman = 0.71 and 0.52 for YPD and YNB, respectively; Figure 3A) and promoter micro-heterogeneity (Spearman = 0.89 and 0.82 for YPD and YNB, respectively; Figure 3B).

As promoter strength and micro-heterogeneity were highly correlated between scRNA-seq and protein expression profiles, we then built a regression model based on **Gaussian processes** to guide promoter selection. Gaussian processes rely on fitting multiple classes of functions that ultimately determine the confidence in the regression model. They do not necessitate any prior knowledge of the mathematical relationship between protein-level sfGFP and scRNA-seq data, making them ideal for the present analysis. Subsequently, we trained the model on experimental data from the 19 YTK promoters (training dataset). As a validation step, we selected six novel promoters (testing dataset) that spanned the entire range of strengths and tested them experimentally using them to control sfGFP expression. Of the selected promoters, four were predicted to be homogeneous (RPL28, CDC19, TDH2, and HCS1) and two were micro- and macro-heterogeneous (HHT1 and HHT2) based on the spread of scRNA-seq read distributions (Figures S7 and S8 in the Supplemental information online). This prediction was corroborated by subsequent flow cytometry analysis of sfGFP expression (Figures S11 and S12 and Tables S5

---

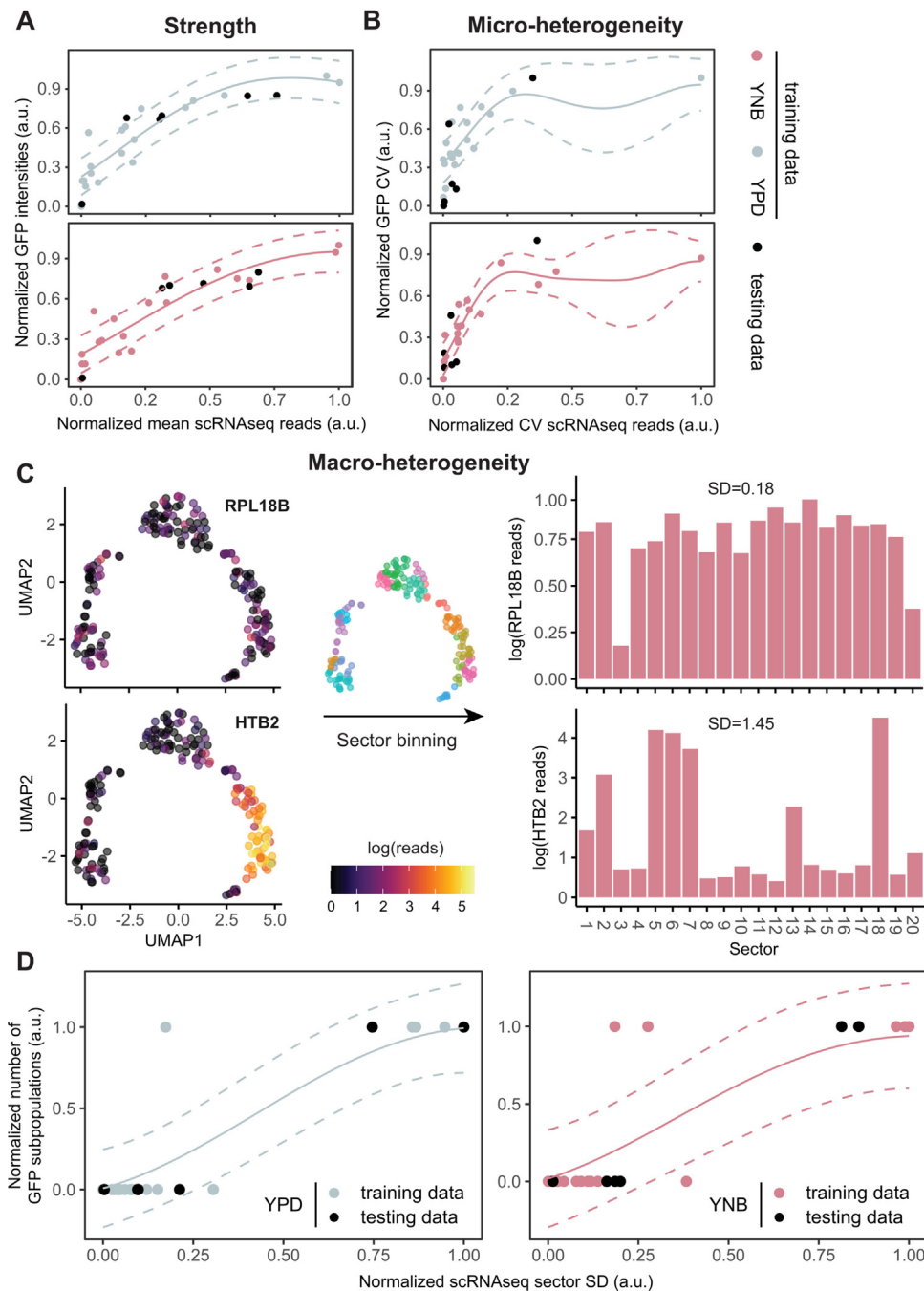
information criterion; pALD6: promoter of aldehyde dehydrogenase 6; pCCW12: promoter of cell wall mannoprotein; pHHF1: promoter of histone H4; pHHF2: promoter of histone H4; pHTB2: promoter of histone H2B; pops: populations; pPAB1: promoter of poly(A)-binding protein; pPGK1: promoter of 3-phosphoglycerate kinase 1; pPOP6: promoter of subunit of RNase MRP; pPSP2: promoter of polymerase suppressor protein 2; pRAD27: promoter of 5' flap endonuclease; pRET2: promoter of subunit of the coatomer complex; pREV1: promoter of DNA-directed DNA polymerase/deoxycytidyl transferase; pRNR1: promoter of large subunit of ribonucleotide-diphosphate reductase; pRNR2: promoter of ribonucleotide-diphosphate reductase 2; pRPL18B: promoter of ribosomal protein L18B; pSAC6: promoter of fimbrin; pTDH3: promoter of glyceraldehyde-3-phosphate dehydrogenase 3; pTEF1: promoter of translation elongation factor 1 $\alpha$ ; pTEF2: promoter of translation elongation factor 1 $\alpha$ ;

(See figure legend at the bottom of the next page.)



**Figure 2. Effect of  $\alpha$ -factor treatment on subpopulation size.** (A) Single-cell sfGFP intensity in exponential-phase cells (grey) and after  $\alpha$ -factor treatment (black). Strains pPGK1, pALD6, and pREV1 were selected due to their homogeneous expression of sfGFP under normal growth conditions. Strains pHHF1, pHHF2, pHTB2, and pRNR1 were selected due to their macro-heterogeneous expression of sfGFP under normal growth conditions. Pop1 is the low-expressing population gate (turquoise); Pop2 is the high-expressing population gate (brown). (B) Proportion of cells in Pop1 and Pop2 in  $\alpha$ -factor-treated (AF) and untreated exponential-phase groups (exp). \*\*\* $P < 0.001$  (Welch's  $t$ -test) for all AF-to-exp comparisons for the same promoter. pALD6: promoter of aldehyde dehydrogenase 6; pHHF1: promoter of histone H4; pHHF2: promoter of histone H4; pHTB2: promoter of histone H2B; pPGK1: promoter of 3-phosphoglycerate kinase 1; pREV1: promoter of DNA-directed DNA polymerase/deoxycytidyl transferase; pRNR1: promoter of large subunit of ribonucleotide-diphosphate reductase; sfGFP: superfolder green fluorescent protein.

and S6 in the Supplemental information online). While the respective sfGFP intensities could be predicted within one standard error (four out of six promoters for YNB and YPD) (Figure 3A), micro-heterogeneity was more difficult to predict accurately, with most promoters outside one SD of the predictions (Figure 3B).



**Figure 3. Prediction of promoter characteristics based on scRNA-seq data.** (A and B) Promoter strength and micro-heterogeneity predictions for sfGFP based on Gaussian process fitting of scRNA-seq data [41] for YPD and YNB log phases. Promoter strengths were log-normalised. The training and testing datasets were min-max normalised for sfGFP and scRNA-seq coefficients of variation and mean intensities. The dashed line represents one standard error of the mean (solid line). (C) Macro-heterogeneity was inferred from the UMAP representation of scRNA-seq data. Single cells were assigned (binned) to a sector based on their UMAP coordinates (coloured UMAP representation in the middle). For a given gene, reads were averaged per sector and the standard deviation (SD) was computed across sectors as a measure

(Figure legend continued at the bottom of the next page.)

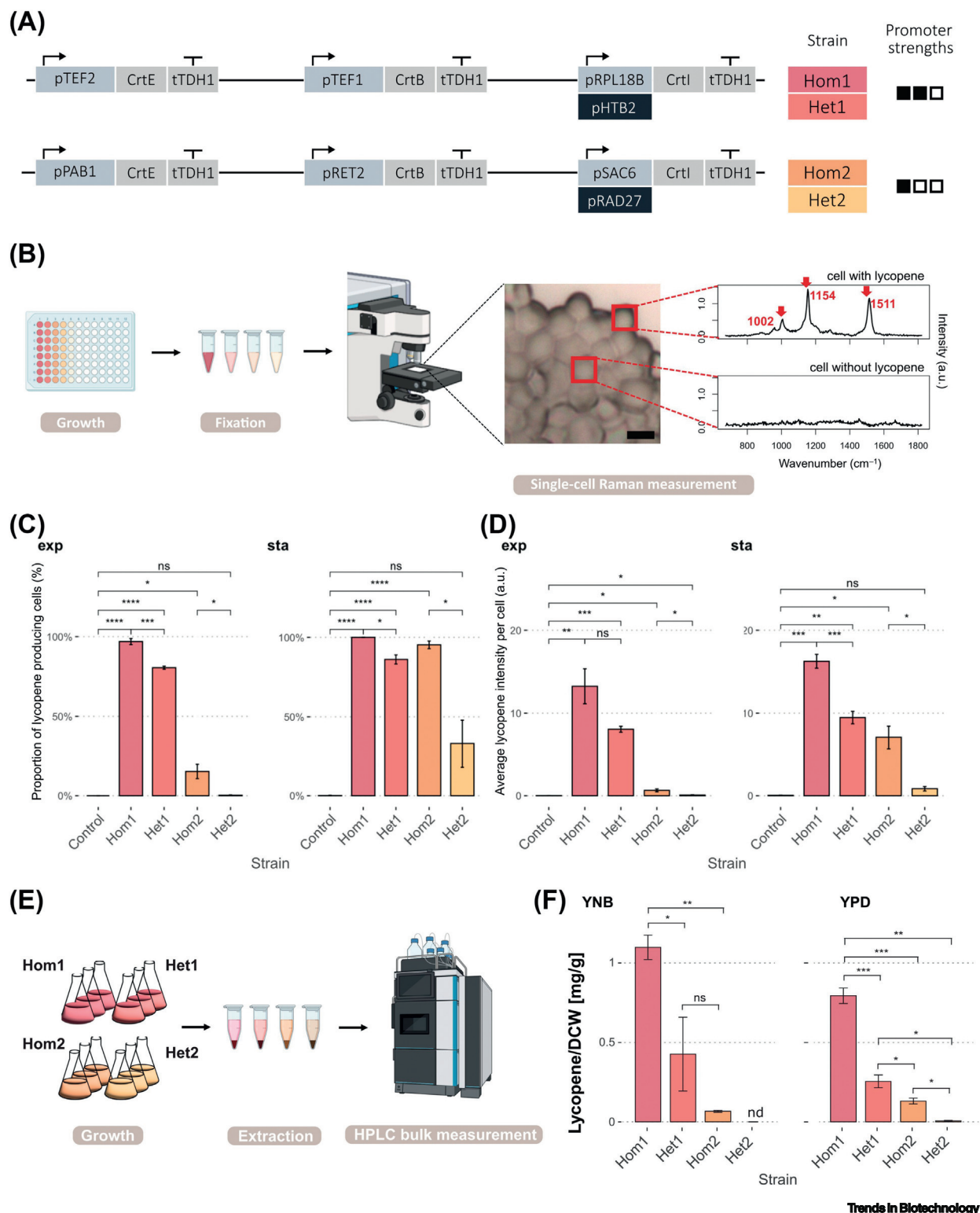
Compared with micro-heterogeneity, macro-heterogeneity was more predictable. Yet our analysis was more challenging due to the sparse representation of scRNA-seq data given the high dropout rates (Figures S7 and S8), and we instead relied on dimensionality reduction. We normalised the dataset and projected it into two dimensions using **uniform manifold approximation and projection (UMAP)**. UMAP allows us to visualise and cluster single cells based on similarities and differences in their transcriptomic profiles. We found that the UMAP representation was fairly informative for predicting macro-heterogeneity. For the 25 promoters tested (19 YTK promoters and six from the validation step), promoters with macro-heterogeneous sfGFP expression appeared clustered on the UMAP representation while homogeneous promoters were almost uniformly expressed across both YPD and YNB datasets (Figures S9 and S10 in the Supplemental information online), and we also observed multimodality in single-cell mRNA read distributions (Figures S7 and S8). This was specifically the case for pHHF1, pHHF2, pHHT1, pHHT2, and pHTB2, with the presence of multiple sfGFP populations and clustered gene expression on the UMAP plot (Figures S1, S9, and S10). To quantify macro-heterogeneity further, we created a macro-heterogeneity score based on the UMAP representation of the scRNA-seq data. In brief, cells were clustered into 20 UMAP ‘sectors’ and the expression of each gene was averaged for each sector (Figure 3C). The SD of averaged gene expression across individual sectors can reflect the multimodality of flow cytometry sfGFP data (Figure 3D). A macro-heterogeneous condition (HTB2) showed stronger local differences, with a higher SD, than a macro-homogeneous condition (RPL18B). For both YPD and YNB datasets, all test data were predicted within one SD by the regression model, indicating that scRNA-seq data are a good predictor of bulk and single-cell promoter characteristics.

#### Promoter-guided metabolic engineering improves yield of lycopene

Next, we asked whether placing metabolic enzymes under homogeneous rather than macro-heterogeneous promoters could improve bioproduction. To answer this question, we built four versions of a lycopene producer strain: Hom1, Het1, Hom2, and Het2 (Figure 4A; Tables S8 and S10–S12 in the Supplemental information online). Lycopene is a valuable carotenoid with applications in food, pharmaceuticals, and cosmetics and can be estimated at the single-cell level using Raman spectroscopy [36,37,43]. The lycopene pathway is composed of three metabolic enzymes: CrtE (geranylgeranyl diphosphate synthase), CrtB (phytoene synthase), and CrtI (phytoene desaturase) (Figure 4A). We observed that using the strongest promoters to express CrtE, CrtB, and CrtI created genetically unstable strains (Figure S13 in the Supplemental information online). As a result, we used medium-strength (Hom1 and Het1) or low-strength promoters (Hom2 and Het2) to express the Crt genes instead. CrtE and CrtB were placed under homogeneous promoters, while CrtI was controlled by either a homogeneous or macro-heterogeneous promoter (Figure 4A). The strengths of individual promoters driving CrtI expression were kept as similar as possible (Figure S14 in the Supplemental information online). Strains were grown in YNB to either the exponential or the stationary phase. Single-cell Raman spectroscopy was conducted to analyse the level of lycopene within each cell in a single population. Due to the resonance effect of the pigment with the incident 532-nm Raman laser, the presence of lycopene could be identified at the single-cell level via three signature Raman bands at 1002, 1154, and

---

of macro-heterogeneity. Here, RPL18B and HTB2 are representative of macro-homogeneous (one population) and macro-heterogeneous (three populations) GFP expression, respectively, during the log phase in YNB. (D) sfGFP macro-heterogeneity prediction based on Gaussian process fitting from scRNA-seq data for YPD and YNB log phases. Promoter strengths were log-normalised. The numbers of sfGFP subpopulations and sector SDs obtained from (C) for both the training and testing datasets were min–max normalised. The dashed line represents one standard error of the mean (solid line). CV: coefficient of variation; HTB2: histone H2B; RPL18B: ribosomal protein L18B; scRNA-seq: single-cell RNA sequencing; sfGFP: superfolder green fluorescent protein; UMAP: uniform manifold approximation and projection; YNB: yeast nitrogen base; YPD: yeast extract/peptone/dextrose.



(See figure legend at the bottom of the next page.)

1511  $\text{cm}^{-1}$ , attributed to C–CH<sub>3</sub> stretching, C–C stretching, and C=C stretching, respectively [43,44]. The concentrations of lycopene were then quantified and compared in single cells and populations (Figure 4B).

The proportion of cells producing lycopene was greater in homogeneous strains (Hom) than in macro-heterogeneous strains (Het) at both the exponential and stationary phases of growth (Figure 4C). Furthermore, average lycopene production was higher in Hom strains than in Het strains, with homogeneous strains producing 1.2–2.9 times more lycopene (Figure 4D; Table S15 in the Supplemental information online). This occurred even when using macro-heterogeneous promoters with stronger expression (Figure S14), as in the case of Hom1 and Het1, in which pHTB2 is stronger than pRPL18B (Figure 4C,D). Overall, the use of homogeneous promoters improved the yield of lycopene.

To further validate the single-cell observations at the population level, bulk lycopene production was quantified using high-performance liquid chromatography (HPLC). The same four strains were grown in both YNB and YPD media for 40 h, followed by pigment extraction and HPLC analysis (Figure 4E). Consistent with Raman measurements, homogeneous promoter strains produced significantly higher lycopene titres than macro-heterogeneous strains under both conditions (Figure 4F; Tables S16–S21 in the Supplemental information online). In YNB, Hom1 and Hom2 accumulated approximately 0.7–2.1  $\text{mg l}^{-1}$  lycopene, whereas Het1 and Het2 produced less than 1  $\text{mg l}^{-1}$  (Table S16). In YPD, homogeneous strains achieved up to 2.5  $\text{mg l}^{-1}$  lycopene, while heterogeneous strains remained below 0.5  $\text{mg l}^{-1}$  (Table S16). Biomass measurements indicated comparable dry cell weights across all strains, with only modest increases in YPD relative to YNB (Tables S18 and S19). When lycopene titres were normalised to dry cell weight, homogeneous strains consistently outperformed heterogeneous strains, producing two to four times more lycopene per gram of biomass (Figure 4F; Table S20). These results confirm that promoter architecture has a strong impact on pathway productivity not only at the single-cell level but also at the population scale.

We also characterised single-cell production using the **FlowRACS** system, which enables high-throughput Raman analysis, and found a consistent overlap between bulk and single-cell production (Figure S18 in the Supplemental information online).

**Figure 4. Strain construction for the production of lycopene and detection by Raman spectroscopy.** (A) Heterologous pathway for lycopene production. Farnesyl pyrophosphate (FPP) is converted into geranylgeranyl diphosphate (GGPP) by CrtE, which is then converted into phytoene and lycopene by CrtB and CrtI, respectively. The constructs were integrated at the URA3 locus to generate Hom1, Het1, Hom2, and Het2 strains. Each strain expresses either homologous (light blue) or macro-heterogeneous (dark blue) levels of CrtI. Promoters were either medium strength (■) or low strength (□), based on flow cytometry measurements of sfGFP (Figure S14). (B) Single-cell lycopene measurement framework. Cells were grown to exponential and stationary phases and fixed in formaldehyde before single-cell Raman spectra were obtained. Lycopene presence can be identified by three characteristic resonance Raman bands at 1002, 1154, and 1511  $\text{cm}^{-1}$ . Average lycopene intensity was measured for each randomly selected cell within a sample, and then an overall average was calculated from more than 200 cells per sample. (C) Proportion of cells producing lycopene in the engineered strains (Hom1, Het1, Hom2, and Het2) during exponential (exp) and stationary (sta) phases. The Control strain did not contain the lycopene cassette (see ‘STAR★Methods’ section). Student’s *t*-test was performed for four biological replicates per strain. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; \*\*\*\**P* < 0.0001; ns: not significant. (D) Average lycopene intensity (a.u.) per cell during exponential (exp) and stationary (sta) phases. Student’s *t*-test was performed for four biological replicates per strain (see ‘STAR★Methods’ section). \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; \*\*\*\**P* < 0.0001. (E) HPLC bulk analysis framework. Cells were grown in a shake flask to stationary phase and lycopene was extracted mechanically. Lycopene solutions were then analysed by HPLC. (F) Quantification of lycopene production relative to dry cell weight from cultures grown in YPD and YNB (Table S20). Welch’s *t*-test was performed for three biological replicates per strain. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; \*\*\*\**P* < 0.0001. ‘nd’ indicates that lycopene was below the detectable range (Table S21). Panels B and E were created using BioRender (<http://biorender.com/>). CrtB: phytoene synthase; CrtI: phytoene desaturase; CrtE: geranylgeranyl diphosphate synthase; DCW: dry cell weight; Het: heterogeneous strain; HPLC: high-performance liquid chromatography; Hom: homogeneous strain; pHTB2: promoter of histone H2B; pPAB1: promoter of poly(A)-binding protein; pRAD27: promoter of 5' flap endonuclease; pRET2: promoter of subunit of the coatomer complex; pRPL18B: promoter of ribosomal protein L18B; pSAC6: promoter of fimbrin; pTEF1: promoter of translation elongation factor 1 $\alpha$ ; pTEF2: promoter of translation elongation factor 1 $\alpha$ ; tTDH1: transcription terminator of glyceraldehyde-3-phosphate dehydrogenase 1; YNB: yeast nitrogen base; YPD: yeast extract/peptone/dextrose; sfGFP: superfolder green fluorescent protein; URA3: orotidine 5'phosphate decarboxylase.

Taken together, these findings demonstrate that promoter homogenisation enables more efficient metabolic production than macro-heterogeneous designs, even when heterogeneous promoters provide higher nominal expression. By reducing the fraction of low- or non-producing cells, homogeneous expression of pathway enzymes maximises yield and robustness, highlighting promoter heterogeneity as a critical design parameter in yeast metabolic engineering.

## Discussion

The presence of phenotypic heterogeneity in clonal populations has been reported across bacterial and yeast expression systems [3–5,31,41,45,46] and has been hypothesised to negatively affect bioproduction [8,9,47]. However, how promoters, as drivers of gene expression, specifically contribute to subpopulation formation and their effect on bioproduction in the context of metabolic engineering is still poorly understood. To assess the impact of promoter heterogeneity on bioproduction, we quantified the YTK promoter system. To measure promoter heterogeneity and strength, we applied Hartigan's dip test and finite mixture modelling to flow cytometry data. We subsequently classified YTK promoters' macro-heterogeneity across the most extensive set of growth conditions to date. We highlighted promoters that are consistently homogeneous, that are consistently macro-heterogeneous, or that transition to macro-heterogeneity upon entry into stationary phase under specific growth conditions (Figure 1B,C; Figures S1 and S2 and Tables S1 and S2). Furthermore, we showed that macro-heterogeneous promoters can form two to three subpopulations with varying levels of expression, and that this macro-heterogeneity is predictable using scRNA-seq data.

Raman spectroscopy was also used to measure single-cell lycopene production in strains expressing either homogeneous or macro-heterogeneous promoters. We demonstrated that avoiding heterogeneous promoters is overall beneficial for metabolic engineering applications, as we could increase lycopene production by up to 2.9-fold using homogeneous promoters compared with macro-heterogeneous promoters of similar strength (Figure 4C,D). Building on prior research on resource competition between subpopulations [2,8,48], we additionally recommend avoiding promoters that form subpopulations to maximise bioproduction.

Phenotypic heterogeneity arising from differences in cell cycle stage has also been identified as a contributor to subpopulation formation in yeast [49]. Similarly, we observed changes in subpopulation ratios when synchronising cell cycles in macro-heterogeneous populations (Figure 2), though we could not completely rescue homogeneity after AF treatment. Moreover, because all constructs were integrated at the same URA3 locus, differences in chromatin accessibility are unlikely to account for the observed differences, suggesting that intrinsic promoter features contribute alongside extrinsic sources of variability. In this context, and consistent with prior work, these results support the view that subpopulation formation arises from a multifactorial mechanism and that improved understanding of heterogeneity in bioprocesses will benefit from models that account for microenvironmental effects [50–53], bet hedging [54], and stochasticity of gene expression [55], alongside practical single-cell measurements.

In addition to promoter heterogeneity and cell cycle stage, our requirement for medium- and low-strength promoters to construct stable cell lines (Figure S13) highlights an important limitation: promoter choice is dictated not only by desired expression levels but also by the metabolic burden imposed on the host. This burden has been observed in other metabolic engineering contexts. For example, in the violacein pathway expressed in bacteria, reducing promoter

strength unexpectedly enhances expression of the Vio genes and increases production of violacein [56]. More broadly, promoter strength is known to influence growth rate, as enhanced heterologous expression can disrupt resource allocation and downstream protein processing [57–59]. While our characterisations enable informed recommendations for promoter selection in *S. cerevisiae*, metabolic burden and overall promoter strength must still be evaluated in conjunction.

Despite these limitations, our findings remain valuable. As modular cloning systems increase in popularity and make microbial bioproduction more accessible, practical, and tuneable, the challenge of promoter heterogeneity persists. By demonstrating that promoter choice should be guided not only by expression strength but also by the ability to sustain a homogeneous cell population across diverse growth conditions, our work provides a framework for improving bioproduction outcomes.

We find that promoter strength inferred from scRNA-seq correlates with protein output at levels comparable to those reported using bulk RNA-seq [60]. At the same time, our single-cell protein measurements could underscore the transcriptional variability observed in scRNA-seq, providing a practical guide for identifying promoters with high or low cell-to-cell variation, enabling more informed promoter selection. By moving beyond profiling intercellular heterogeneity in previous scRNA-seq studies, our approach demonstrates clear potential to improve downstream outcomes in metabolic engineering applications. With recent advances in bacterial scRNA-seq [61], we believe this methodology could also be extended to bacterial hosts and even novel model organisms. Furthermore, as promoter engineering continues to advance and the collection of synthetic promoters within available toolkits expands [22,62], future characterisations of novel promoters should also consider the impact of promoter heterogeneity in different growth conditions to maximise strain productivity.

## Concluding remarks

This work presents a framework for identifying macro-heterogeneous promoters by applying statistical models to flow cytometry and scRNA-seq data. In doing so, we were able to identify constitutive promoters which behave heterogeneously in multiple growth conditions. Additionally, by avoiding macro-heterogeneous promoters in a three-step metabolic pathway, we were able to demonstrate a threefold increase in the single-cell bioproduction of lycopene in *S. cerevisiae*. Unlike previous studies on population-level expression heterogeneity, our proof-of-concept results confirm that promoter heterogeneity impacts bioproduction at both the single-cell and bulk levels in engineered pathways. Future efforts in metabolic pathway engineering should consider promoter macro-heterogeneity as an important design parameter and investigate whether its effects become amplified in more complex pathways and under industrial-scale conditions (see [Outstanding questions](#)). Nevertheless, by demonstrating the consequences of promoter macro-heterogeneity, our findings highlight the importance of homogenising promoters before pathway construction to improve yield.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- Experimental model and study participant details
  - Bacterial strains
  - Yeast strains
- Strain design
  - Golden Gate assembly
  - Construction of lycopene strains

## Outstanding questions

Do more complex metabolic pathways experience the same level of detrimental effects from use of macro-heterogeneous promoters?

Do the detrimental effects of macro-heterogeneous promoters become amplified during large-scale industrial bioreactor fermentations?

Is promoter heterogeneity short-lived in individual cells, or does it persist? If so, could a stable, low-producing population confer a growth advantage during longer fermentations and thereby increase bioproduction?

Can scRNA-seq predict promoter strength and cell-to-cell heterogeneity across organisms and conditions, and can these predictions guide the future design of promoters tuned for specific applications?

In addition to cell cycle state, what combination of intrinsic and extrinsic factors governs promoter heterogeneity, and can a homogeneous expression state be reliably restored from macro-heterogeneous promoters?

- Yeast transformation
- Colony PCR
- Flow cytometry of pYTK strains
- Subpopulation analysis of strains
- Analysis of growth arrested cells
  - Flow cytometry
  - Data analysis
- Cell cycle microscopy analysis
- Protein level prediction from scRNA-seq data
- Sample preparation for Raman spectroscopy
- Single-cell lycopene measurements using Raman spectroscopy
- Lycopene quantification using HPLC and FlowRACS

## RESOURCE AVAILABILITY

### Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Rodrigo Ledesma-Amaro ([r.ledesma-amaro@imperial.ac.uk](mailto:r.ledesma-amaro@imperial.ac.uk)).

### Materials availability

All key plasmids and engineered *S. cerevisiae* strains generated in this study are listed in Table S8 and are available from the lead contact upon reasonable request.

### Data and code availability

Source code for statistical modelling is available on GitHub, as indicated in the Key Resources Table, and is publicly available as of the date of publication. Flow cytometry data can be made available upon request from the lead contact. Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

## Author contributions

R.L.A., D.P., C.K., and K.S. conceived and designed the experiments. D.P. generated the strains, performed flow cytometry, data analysis, and Raman spectroscopy sample collection. K.S. conducted scRNA-seq data analysis and designed constructs for model validation. C.K. performed HPLC, microscopy, and FlowRACS, along with data analysis. J.X. performed single-cell Raman spectroscopy experiments and analysed the resulting data under the supervision of W.E.H. R.L.A. supervised all work involved. D.P., K.S., C.K., and J.X. wrote the manuscript, with all authors contributing to the edits of the final draft.

## Acknowledgements

R.L.A. received funding from the Biotechnology and Biological Sciences Research Council (BB/R01602X/1, BB/T013176/1, BB/T011408/1—19-ERACoBioTech-33 SyCoLim, BB/X01911X/1, and BB/Y008510/1—Engineering Biology Hub for Microbial Foods), Engineering and Physical Sciences Research Council (AI-4-EB BB/W013770/1 and EEBio Programme Grant EP/Y014073/1), Yeast4Bio COST Action 18229, the European Research Council (DEUSBIO – 949080), and the Bio-based Industries Joint Undertaking (PERFECOAT – 101022370) under the European Union's Horizon 2020 research and innovation programme, the European Innovation Council under grant agreement No. 101098826 (SKINDEV), the Bezos Centre for Sustainable Proteins at Imperial (BCSP/IC/001), and the Imperial College London UKRI Impact Acceleration Account (EPSRC – EP/X52556X/1 and BBSRC – BB/X511055/1). K.S. acknowledges a postdoctoral fellowship from the European Molecular Biology Organization (ALTF 769-2021) and a UKRI-Marie Skłodowska-Curie Actions Postdoctoral Fellowship (UNICOH). W.E.H. thanks EPSRC (EP/M02833X/1) for Raman instrument support. The authors would also like to thank Dr Will Shaw (Tom Ellis's lab) for his assistance in constructing the pYTK-sfGFP plasmids.

## Declaration of interests

The authors declare no competing interests.

## Supplemental information

Supplementary information associated with this article can be found online at <https://doi.org/10.1016/j.tibtech.2026.08.002>.

## References

- Del Vecchio, D. *et al.* (2016) Control theory meets synthetic biology. *J. R. Soc. Interface* 13, 20160380
- Kamrad, S. *et al.* (2023) Metabolic heterogeneity and cross-feeding within isogenic yeast populations captured by DILAC. *Nat. Microbiol.* 8, 441–454
- Newman, J.R.S. *et al.* (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441, 840–846
- Saint, M. *et al.* (2019) Single-cell imaging and RNA sequencing reveal patterns of gene expression heterogeneity during fission yeast growth and adaptation. *Nat. Microbiol.* 4, 480–491
- Nadal-Ribelles, M. *et al.* (2019) Sensitive high-throughput single-cell RNA-seq reveals within-clonal transcript correlations in yeast populations. *Nat. Microbiol.* 4, 683–692
- Takhaviev, V. and Heinemann, M. (2018) Metabolic heterogeneity in clonal microbial populations. *Curr. Opin. Microbiol.* 45, 30–38
- Lee, K.-S. *et al.* (2013) Characterization of *Saccharomyces cerevisiae* promoters for heterologous gene expression in *Kluyveromyces marxianus*. *Appl. Microbiol. Biotechnol.* 97, 2029–2041
- Binder, D. *et al.* (2017) Homogenizing bacterial cell factories: analysis and engineering of phenotypic heterogeneity. *Metab. Eng.* 42, 145–156
- Delvigne, F. *et al.* (2014) Metabolic variability in bioprocessing: implications of microbial phenotypic heterogeneity. *Trends Biotechnol.* 32, 608–616
- Raser, J.M. and O'Shea, E.K. (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304, 1811–1814
- Liu, J. *et al.* (2018) Bimodality of gene expression from yeast promoter can be instigated by DNA context, inducing conditions and strain background. *FEMS Yeast Res.* 18, foy047
- Blake, W.J. *et al.* (2006) Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* 24, 853–865
- Lubliner, S. *et al.* (2015) Core promoter sequence in yeast is a major determinant of expression level. *Genome Res.* 25, 1008–1017
- Peng, B. *et al.* (2015) Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. *Microb. Cell Factories* 14, 91
- Galdieri, L. *et al.* (2010) Transcriptional regulation in yeast during diauxic shift and stationary phase. *OMICS* 14, 629–638
- Seeger, M. *et al.* (2023) A dynamical stochastic model of yeast translation across the cell cycle. *Heliyon* 9, e13101
- Bisschops, M.M.M. *et al.* (2017) Extreme calorie restriction in yeast retentostats induces uniform non-quiescent growth arrest. *Biochim. Biophys. Acta* 1864, 231–242
- Spellman, P.T. *et al.* (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* 9, 3273–3297
- Kelliher, C.M. *et al.* (2018) Layers of regulation of cell-cycle gene expression in the budding yeast *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 29, 2644–2655
- Sharon, E. *et al.* (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* 30, 521–530
- Carey, L.B. *et al.* (2013) Promoter sequence determines the relationship between expression level and noise. *PLoS Biol.* 11, e1001528
- Kotopka, B.J. and Smolke, C.D. (2020) Model-driven generation of artificial yeast promoters. *Nat. Commun.* 11, 2113
- Yu, T.C. *et al.* (2021) Multiplexed characterization of rationally designed promoter architectures deconstructs combinatorial logic for IPTG-inducible systems. *Nat. Commun.* 12, 325
- Sharon, E. *et al.* (2014) Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome Res.* 24, 1698–1706
- Curran, K.A. *et al.* (2014) Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat. Commun.* 5, 4002
- Nadal-Ribelles, M. *et al.* (2024) The rise of single-cell transcriptomics in yeast. *Yeast* 41, 158–170
- Brettner, L. *et al.* (2024) An ultra high-throughput, massively multiplexable, single-cell RNA-seq platform in yeasts. *Yeast* 41, 242–255
- Gasch, A.P. *et al.* (2017) Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. *PLoS Biol.* 15, e2004050
- Jariani, A. *et al.* (2020) A new protocol for single-cell RNA-seq reveals stochastic gene expression during lag phase in budding yeast. *Elife* 9, e55320
- Nadal-Ribelles, M. *et al.* (2025) Transcriptional heterogeneity shapes stress-adaptive responses in yeast. *Nat. Commun.* 16, 2631
- Sampaio, N.M.V. *et al.* (2022) Dynamic gene expression and growth underlie cell-to-cell heterogeneity in *Escherichia coli* stress response. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2115032119
- Khlebnikov, A. *et al.* (2002) Modulation of gene expression from the arabinose-inducible *araBAD* promoter. *J. Ind. Microbiol. Biotechnol.* 29, 34–37
- Lee, M.E. *et al.* (2015) A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* 4, 975–986
- Shaw, W.M. *et al.* (2022) Inducible expression of large gRNA arrays for multiplexed CRISPRai applications. *Nat. Commun.* 13, 4984
- Kobayashi-Kirschvink, K.J. *et al.* (2024) Prediction of single-cell RNA expression profiles in live cells by Raman microscopy with Raman2RNA. *Nat. Biotechnol.* 42, 1726–1734
- Song, Y. *et al.* (2017) Single-cell genomics based on Raman sorting reveals novel carotenoid-containing bacteria in the Red Sea. *Microb. Biotechnol.* 10, 125–137
- Borek-Dorosz, A. *et al.* (2022) Raman-based spectrophototyping of the most important cells of the immune system. *J. Adv. Res.* 41, 191–203
- Müller, M. *et al.* (2012) A prototrophic deletion mutant collection for yeast metabolomics and systems biology. *Nat. Biotechnol.* 30, 1176–1178
- Leisch, F. (2004) FlexMix: a general framework for finite mixture models and latent class regression in R. *J. Stat. Softw.* 11, 1–18
- Lipke, P.N. *et al.* (1976) Morphogenic effects of alpha-factor on *Saccharomyces cerevisiae* a cells. *J. Bacteriol.* 127, 610–618
- Jackson, C.A. *et al.* (2020) Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. *Elife* 9, e51254
- Shabestary, K. *et al.* (2024) Phenotypic heterogeneity follows a growth-viability tradeoff in response to amino acid identity. *Nat. Commun.* 15, 6515
- Li, M. *et al.* (2012) Rapid resonance Raman microspectroscopy to probe carbon dioxide fixation by single cells in microbial communities. *ISME J.* 6, 875–885
- Hoskins, L.C. (1984) Resonance Raman spectroscopy of beta-carotene and lycopene, a physical chemistry experiment. *J. Chem. Educ.* 61, 460
- Avery, S.V. (2006) Microbial cell individuality and the underlying sources of heterogeneity. *Nat. Rev. Microbiol.* 4, 577–587
- Calabrese, F. *et al.* (2019) Quantitation and comparison of phenotypic heterogeneity among single cells of monoclonal microbial populations. *Front. Microbiol.* 10, 2814
- Fernandes, R.L. *et al.* (2011) Experimental methods and modeling techniques for description of cell population heterogeneity. *Biotechnol. Adv.* 29, 575–599
- Wang, J. *et al.* (2015) Natural variation in preparation for nutrient depletion reveals a cost-benefit tradeoff. *PLoS Biol.* 13, e1002041
- Santos, A. *et al.* (2015) Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Res.* 43, D1140–D1144
- Acar, M. *et al.* (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* 40, 471–475
- Müller, S. *et al.* (2010) Origin and analysis of microbial population heterogeneity in bioprocesses. *Curr. Opin. Biotechnol.* 21, 100–113
- Lara, A.R. *et al.* (2006) Living with heterogeneities in bioreactors: understanding the effects of environmental gradients on cells. *Mol. Biotechnol.* 34, 355–381
- Botman, D. *et al.* (2020) An improved ATP FRET sensor for yeast shows heterogeneity during nutrient transitions. *ACS Sens.* 5, 814–822

54. Davidson, G.S. *et al.* (2011) The proteomics of quiescent and nonquiescent cell differentiation in yeast stationary-phase cultures. *Mol. Biol. Cell* 22, 988–998
55. Veening, J.-W. *et al.* (2008) Bistability, epigenetics, and bet-hedging in bacteria. *Ann. Rev. Microbiol.* 62, 193–210
56. Jones, J.A. *et al.* (2015) ePathOptimize: a combinatorial approach for transcriptional balancing of metabolic pathways. *Sci. Rep.* 5, 11301
57. Bienick, M.S. *et al.* (2014) The interrelationship between promoter strength, gene expression, and growth rate. *PLoS One* 9, e109105
58. Rugbjerg, P. and Sommer, M.O.A. (2019) Overcoming genetic heterogeneity in industrial fermentations. *Nat. Biotechnol.* 37, 869–876
59. Kastberg, L.L.B. *et al.* (2022) Burden imposed by heterologous protein production in two major industrial yeast cell factories: identifying sources and mitigation strategies. *Front. Fungal Biol.* 3, 827704
60. Csárdi, G. *et al.* (2015) Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. *PLoS Genet.* 11, e1005206
61. Ma, P. *et al.* (2023) Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states. *Cell* 186, 877–891.e14
62. Tang, H. *et al.* (2020) Promoter architecture and promoter engineering in *Saccharomyces cerevisiae*. *Metabolites* 10, 320
63. McCarthy, D.J. *et al.* (2017) Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33, 1179–1186
64. Schindelin, J. *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682
65. Shaw, W.M. *et al.* (2019) Engineering a model cell for rational tuning of GPCR signaling. *Cell* 177, 782–796.e27
66. Wang, X. *et al.* (2023) Robust spontaneous Raman flow cytometry for single-cell metabolic phenome profiling via pDEP-DLD-RFC. *Adv. Sci. (Weinh.)* 10, e2207497
67. Wang, X. *et al.* (2025) Label-free high-throughput live-cell sorting of genome-wide random mutagenesis libraries for metabolic traits by Raman flow cytometry. *Proc. Natl. Acad. Sci. U. S. A.* 122, e2503641122

## STAR★METHODS

## KEY RESOURCES TABLE

| Reagent or resource                                                             | Source     | Identifier                                                                                                              |
|---------------------------------------------------------------------------------|------------|-------------------------------------------------------------------------------------------------------------------------|
| Deposited data                                                                  |            |                                                                                                                         |
| Platecurver                                                                     | This study | <a href="https://github.com/davinap/YTK-heterogeneity-platecurver">github.com/davinap/YTK-heterogeneity-platecurver</a> |
| Plate reader GFP intensity analysis                                             | This study | <a href="https://github.com/davinap/YTK-heterogeneity-GFP">github.com/davinap/YTK-heterogeneity-GFP</a>                 |
| Flow cytometry analysis                                                         | This study | <a href="https://github.com/davinap/YTK-heterogeneity-flow">github.com/davinap/YTK-heterogeneity-flow</a>               |
| Flow cytometry subpopulation analysis                                           | This study | <a href="https://github.com/davinap/YTK-heterogeneity-analysis">github.com/davinap/YTK-heterogeneity-analysis</a>       |
| Flow cytometry subpopulation analysis on model selected promoters               | This study | <a href="https://github.com/davinap/YTK-heterogeneity-model">github.com/davinap/YTK-heterogeneity-model</a>             |
| Flow cytometry analysis of growth arrested cells                                | This study | <a href="https://github.com/davinap/YTK-heterogeneity-GA">github.com/davinap/YTK-heterogeneity-GA</a>                   |
| scRNAseq analysis                                                               | This study | <a href="https://github.com/KiyanShabestary/YTK-heterogeneity">http://github.com/KiyanShabestary/YTK-heterogeneity</a>  |
| Experimental models: Organisms/strains                                          |            |                                                                                                                         |
| Sequence maps for the pYTK-sfGFP experimental strains are available in Table S8 | This study | N/A                                                                                                                     |
| Sequence maps for the lycopene strains are available in Table S12               | This study | N/A                                                                                                                     |
| pTDH3                                                                           | This study | S2431                                                                                                                   |
| pCCW12                                                                          | This study | S2432                                                                                                                   |
| pPGK1                                                                           | This study | S2433                                                                                                                   |
| pHHF2                                                                           | This study | S2434                                                                                                                   |
| pTEF1                                                                           | This study | S2435                                                                                                                   |
| pTEF2                                                                           | This study | S2436                                                                                                                   |
| pHHF1                                                                           | This study | S2437                                                                                                                   |
| pHTB2                                                                           | This study | S2438                                                                                                                   |
| pRPL18B                                                                         | This study | S2439                                                                                                                   |
| pALD6                                                                           | This study | S2440                                                                                                                   |
| pPAB1                                                                           | This study | S2441                                                                                                                   |
| pRET2                                                                           | This study | S2442                                                                                                                   |
| pRNR1                                                                           | This study | S2443                                                                                                                   |
| pSAC6                                                                           | This study | S2444                                                                                                                   |
| pRNR2                                                                           | This study | S2445                                                                                                                   |
| pPOP6                                                                           | This study | S2446                                                                                                                   |
| pRAD27                                                                          | This study | S2447                                                                                                                   |
| pPSP2                                                                           | This study | S2448                                                                                                                   |
| pREV1                                                                           | This study | S2449                                                                                                                   |
| Control                                                                         | This study | S2450                                                                                                                   |
| pHCS1                                                                           | This study | S2477                                                                                                                   |
| pHHT1                                                                           | This study | S2478                                                                                                                   |
| pTDH2                                                                           | This study | S2479                                                                                                                   |
| pHHT2                                                                           | This study | S2480                                                                                                                   |
| pCDC19                                                                          | This study | S2481                                                                                                                   |
| pRPL28                                                                          | This study | S2482                                                                                                                   |

(continued on next page)

(continued)

| Reagent or resource                                                     | Source                 | Identifier |
|-------------------------------------------------------------------------|------------------------|------------|
| Hom1                                                                    | This study             | S3336      |
| Hom2                                                                    | This study             | S3337      |
| Het1                                                                    | This study             | S3332      |
| Het2                                                                    | This study             | S3334      |
| Oligonucleotides                                                        |                        |            |
| URA3 barcode forward                                                    | Lee <i>et al.</i> [33] | N/A        |
| URA3 barcode reverse                                                    | Lee <i>et al.</i> [33] | N/A        |
| Recombinant DNA                                                         |                        |            |
| Plasmids maps for building pYTK-sfGFP strains are available in Table S7 | This study             | N/A        |
| Plasmids maps for building lycopene strains are available in Table S10  | This study             | N/A        |
| pYTK096                                                                 | Addgene                | #65203     |
| pYTK001                                                                 | Addgene                | #65108     |
| pTDH3 promoter                                                          | Addgene                | #65116     |
| pCCW12 promoter                                                         | Addgene                | #65117     |
| pPGK1 promoter                                                          | Addgene                | #65118     |
| pHHF2 promoter                                                          | Addgene                | #65119     |
| pTEF1 promoter                                                          | Addgene                | #65120     |
| pTEF2 promoter                                                          | Addgene                | #65121     |
| pHHF1 promoter                                                          | Addgene                | #65122     |
| pHTB2 promoter                                                          | Addgene                | #65123     |
| pRPL18B promoter                                                        | Addgene                | #65124     |
| pALD6 promoter                                                          | Addgene                | #65125     |
| pPAB1 promoter                                                          | Addgene                | #65126     |
| pRET2 promoter                                                          | Addgene                | #65127     |
| pRNR1 promoter                                                          | Addgene                | #65128     |
| pSAC6 promoter                                                          | Addgene                | #65129     |
| pRNR2 promoter                                                          | Addgene                | #65130     |
| pPOP6 promoter                                                          | Addgene                | #65131     |
| pRAD27 promoter                                                         | Addgene                | #65132     |
| pPSP2 promoter                                                          | Addgene                | #65133     |
| pREV1 promoter                                                          | Addgene                | #65134     |
| pTDH1 terminator                                                        | Addgene                | #65163     |
| sfGFP                                                                   | Addgene                | #123056    |
| CrtE                                                                    | This study             | Table S11  |
| CrtB                                                                    | This study             | Table S11  |
| CrtI                                                                    | This study             | Table S11  |
| pWS041                                                                  | Addgene                |            |
| pWS043                                                                  | Addgene                |            |
| pWS044                                                                  | Addgene                |            |
| pHLM minichromosome                                                     | Addgene                | #64168     |
| pHLUM minichromosome                                                    | Addgene                | #40276     |

(continued)

| Reagent or resource                                               | Source                   | Identifier                                                                                                                                                                      |
|-------------------------------------------------------------------|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Software and algorithms                                           |                          |                                                                                                                                                                                 |
| Platecurver                                                       | This study               | <a href="https://github.com/davinap/YTK-heterogeneity-platecurver">github.com/davinap/YTK-heterogeneity-platecurver</a>                                                         |
| Plate reader GFP intensity analysis                               | This study               | <a href="https://github.com/davinap/YTK-heterogeneity-GFP">github.com/davinap/YTK-heterogeneity-GFP</a>                                                                         |
| Flow cytometry analysis                                           | This study               | <a href="https://github.com/davinap/YTK-heterogeneity-flow">github.com/davinap/YTK-heterogeneity-flow</a>                                                                       |
| Flow cytometry subpopulation analysis                             | This study               | <a href="https://github.com/davinap/YTK-heterogeneity-analysis">github.com/davinap/YTK-heterogeneity-analysis</a>                                                               |
| Flow cytometry subpopulation analysis on model selected promoters | This study               | <a href="https://github.com/davinap/YTK-heterogeneity-model">github.com/davinap/YTK-heterogeneity-model</a>                                                                     |
| Flow cytometry analysis of growth arrested cells                  | This study               | <a href="https://github.com/davinap/YTK-heterogeneity-GA">github.com/davinap/YTK-heterogeneity-GA</a>                                                                           |
| scRNAseq analysis                                                 | This study               | <a href="https://github.com/KiyanShabestary/YTK-heterogeneity">http://github.com/KiyanShabestary/YTK-heterogeneity</a>                                                          |
| Flexmix package                                                   | [39]                     | <a href="https://jef.works/blog/2017/08/05/a-practical-introduction-to-finite-mixture-models/">jef.works/blog/2017/08/05/a-practical-introduction-to-finite-mixture-models/</a> |
| Hartigan's dip test                                               |                          | <a href="https://github.com/mmaechler/diptest">github.com/mmaechler/diptest</a>                                                                                                 |
| Scuttle package                                                   | [63]                     |                                                                                                                                                                                 |
| GauPro package                                                    |                          | <a href="https://github.com/CollinErickson/GauPro">github.com/CollinErickson/GauPro</a>                                                                                         |
| ImageJ/Fiji                                                       | [64]                     | <a href="https://imagej.net/">https://imagej.net/</a> RRID: SCR_003070                                                                                                          |
| Chromeleon v7                                                     | ThermoFisher             | Version 7                                                                                                                                                                       |
| RStudio                                                           |                          | <a href="https://cran.rstudio.com/">https://cran.rstudio.com/</a>                                                                                                               |
| FlowJo™                                                           | BD                       | v10.4                                                                                                                                                                           |
| Other                                                             |                          |                                                                                                                                                                                 |
| 96-well plate                                                     | Corning                  | 3596                                                                                                                                                                            |
| Breathe-easy film                                                 | Sigma-Aldrich            | Z380059                                                                                                                                                                         |
| Shaking incubator                                                 | Infors                   | Multitron                                                                                                                                                                       |
| Plate reader                                                      | Biotek                   | Synergy                                                                                                                                                                         |
| Plate reader                                                      | Byonoy                   | Enzo Absorbance 96                                                                                                                                                              |
| Flow cytometer                                                    | Invitrogen               | Attune Nxt                                                                                                                                                                      |
| Microscope                                                        | Nikon                    | Eclipse Ti-E                                                                                                                                                                    |
| Camera                                                            | Hamamatsu                | ORCA-Flash4 sCMOS                                                                                                                                                               |
| Light source                                                      | CoolLED                  | PE-300 LED                                                                                                                                                                      |
| Raman microscopic spectrometer                                    | HR Evolution             | N/A                                                                                                                                                                             |
| Tissue homogeniser                                                | Bertin                   | Precellys                                                                                                                                                                       |
| Vanquish KITT HPLC system                                         | Thermo Fisher Scientific |                                                                                                                                                                                 |

## Experimental model and study participant details

### Bacterial strains

All cloning steps were undertaken in Turbo *E. coli* competent cells (NEB C2984). Cells were cultured in LB medium (Merck 28713) with shaking at 250 rpm or plated on LB agar (Merck 1.10283.0500), each supplemented with the appropriate antibiotic - 50 µg/ml of kanamycin (Sigma-Aldrich K1377) or 25 µg/ml chloramphenicol (Acros Organics 22792026). Cultures were incubated at 37 °C.

### Yeast strains

All *S. cerevisiae* strains used in this study are available in the Key Resources Table and Supplementary Document S1. All strains were derived from the BY4741 strain (ATCC 4040002), and routinely cultured in yeast extract peptone dextrose (YPD) medium (20 g/L peptone, Merck, Cat. No. 1.11931.1000; 10 g/L yeast extract, Sigma-Aldrich, Cat. No. Y1625; 2% (w/v) glucose, VWR, Cat. No. 101174Y) at 30 °C with shaking at 250 rpm. Transformants were selected by auxotrophy at 30 °C either on synthetic dropout agar plates lacking histidine, leucine and uracil (SC-HLU-) (6.7 g/L yeast nitrogen base, Sigma-Aldrich Y0626; 1.4 g/L dropout

supplement, Sigma Y2001; 76 mg tryptophan, Sigma-Aldrich T0254; 20 g/L agar, VWR 20767; 2 % sterile-filtered glucose, VWR 101174Y), or on YNB plates (6.7 g/L yeast nitrogen base, Sigma-Aldrich Y0626; 20 g/L agar, VWR 20767; 2 % sterile-filtered glucose, VWR 101174Y). Experimental cultures were incubated with shaking at 30 °C in either YPD, YPD10 (20 g/L peptone; 10 g/L yeast extract; 10 % glucose), YNB (6.7 g/L yeast nitrogen base, Sigma-Aldrich Y0626; 2 % sterile-filtered glucose), or YEPG (20 g/L peptone; 10 g/L yeast extract; 1 % glycerol, VWR 24388.295; 1 % ethanol, VWR 20821.330), as indicated in the corresponding methods.

## Strain design

### Golden Gate assembly

Plasmids were constructed using the Golden Gate Assembly method, as outlined in [33]. Each pYTK-sfGFP-tTDH1 transcriptional unit was integrated into pYTK096 in a Bsal Golden Gate reaction containing 50 fmol of each Level 0 part, 0.2 µl of Bsal (NEB R3733L), 0.5 µl T7 DNA ligase (NEB, M0318L), 1 µl of 10X T4 buffer (NEB, B0202S) and nuclease free water to a total of 10 µl. The resulting pYTK-sfGFP constructs used have been adapted from ref [65], and are specified in Table S7. They contain flanking regions for site-directed integration in the URA3 gene locus of the yeast genome. The reaction was cycled with the following settings: 25 cycles of 37 °C for 2 min and 16 °C for 5 min; 60 °C for 10 min; 80 °C for 10 min. 10 µl of the resulting reaction was added to Turbo *E. coli* competent cells (NEB C2984), incubated for 10 min on ice, heat-shocked at 42 °C for 1 min and placed in ice for 1 min. 100 µl of LB (VWR 84649.0500) was added before recovery at 37 °C with shaking for 15 min. Cells were plated on LB agar containing 50 µg/ml of kanamycin. White colonies, containing the right insert, were selected and grown overnight in 5 ml LB media containing 50 µg/ml kanamycin. Plasmid DNA was isolated using a Qiagen Miniprep kit (Qiagen, 27104) according to the manufacturer's instructions. Plasmids were verified by restriction digestion with NotI (NEB, R3189L) using gel electrophoresis (0.8 %). Final constructs are available in Table S8.

### Construction of lycopene strains

CrtE, CrtB and CrtI genes (Table S11) were integrated into pYTK001 in a Golden Gate reaction as follows: 50 fmol of either CrtE, CrtB or CrtI; 0.5 µl of BsmBI (NEB, R0580L); 0.5 µl T7 DNA Ligase; 1 µl 10X T4 buffer; 0.5 µl of 50 fmol pYTK001 (Addgene, 65108); nuclease-free water up to 10 µl. Reactions were cycled: 25 times at 37 °C for 2 min and 16 °C for 5 min; 60 °C for 10 min; 80 °C for 10 min. Turbo competent *E. coli* were then transformed and plasmid DNA was purified. For Level 1 construction, the parts listed in Tables S10 and S11 were combined in a Bsal Golden Gate reaction, as outlined in the previous section. Promoters for the medium-strength strain and low-strength strains were chosen because they had similar levels of expression in flow cytometry (Figure S14). For multigene Level 2 construction, the parts in Table S12 were combined in a BsmBI Golden Gate reaction with cycle settings as outlined previously. The final Hom1, Het1, Hom2, Het2 constructs are available in Table S8. After *E. coli* transformation and miniprep, 500 ng of each plasmid was digested with NotI, and integrated into the URA3 locus as described below. Cells were saved as glycerol stocks at -80 °C until use.

## Yeast transformation

500 ng of each Level 1 plasmid was digested in a reaction containing: 0.3 µl NotI, 1 µl CutSmart buffer (NEB B6004S), and nuclease-free water up to a total of 10 µl. The following settings were used: 37 °C for 12 h, 65 °C for 20 min. 1 colony of BY4741 (ATCC 4040002) was inoculated into 3 ml of YPD and grown overnight at 30 °C with shaking at 250 rpm. The next day, 100 µl of overnight culture was added to 5 ml of YPD and incubated with shaking at 30 °C for 5 h. Cells were spun down at 5000 g for 5 min (Eppendorf, 5424 R), washed with 5 ml of ddH<sub>2</sub>O and spun down at 5000 g for 5 min. The water was removed and cells were resuspended in 300 µl of 0.1 M lithium acetate (LiOAc) (Sigma-Aldrich, 920320). 100 µl of cells were added to 1.5 ml tubes. After 10 min, cells were spun down at 5000g for 5 min. The supernatant was removed and the cell pellet was gently resuspended in a DNA cocktail containing: 10 µl of NotI digested plasmid, 300 ng of pHLM minichromosome (Addgene, 64168), 10 µl of boiled salmon sperm (Invitrogen, 15632011), and 42.5 µl of nuclease-free water. For Control cells, NotI digested plasmid was omitted and cells were instead transformed with pHLUM minichromosome (Addgene, 40276). A polyethylene glycol-LiOAc mix containing 260 µl of PEG-3350 (Sigma-Aldrich, 202444) and 36 µl 1 M LiOAc was then added. Cells were gently aspirated up and down to homogenise the reaction mix before incubation at 42 °C for 25 min in a heat block (Fisherbrand, 15397928). Cells were centrifuged at 5000 g for 1 min (Eppendorf 5425R), resuspended in 100 µl of ddH<sub>2</sub>O and spread on SC-HLU- plates. Methionine was included in the plates to reduce burden on transformants. Plates were incubated at 30 °C for 2 days then 3 colonies from each transformation were re-streaked on YNB plates to remove cells without pHLM and reduce background plasmid on the plates before colony PCR. Plates were incubated at 30 °C for 2 days. Strains (Table S8) can be provided upon request.

### Colony PCR

To check for integration of the transcription unit at the URA3 gene locus, colony PCR was performed as per the manufacturer's protocol for yeast using the Phire Plant Direct PCR Master Mix (Thermo Scientific, F160L). The URA3 primers used are listed in Table S4 in ref [33]. PCR products were checked on an agarose gel (0.8 %). Following the confirmation of 500 bp bands at the 5' and 3' end, positive colonies were inoculated in 3 ml of YNB media and grown for 48 h at 30 °C with shaking at 250 rpm. Cultures were stored in 50 % v/v glycerol (VWR 24388.295) at -80 °C until required.

### Flow cytometry of pYTK strains

The following media were prepared in advance: YPD, YPD10, YNB, and YEPG. For pre-culturing, 125 µl YPD was dispensed into each well of a 96-well plate (Corning 3596). Yeast strains were inoculated by transferring approximately 1 mm of frozen glycerol stock into each well using a sterile pipette tip. Plates were sealed with gas-permeable film (Breathe-Easy, Sigma-Aldrich Z380059) and incubated for 48 h at 30 °C with orbital shaking at 700 rpm (Infors Multitron). Cells were harvested by centrifugation at 3000 × g for 2 min (Eppendorf 5810), washed three times with sterile PBS (Gibco 10010023), and finally resuspended in 125 µl PBS. For experimental cultures, 125 µl of fresh medium (YPD, YPD10, YNB, or YEPG) was added to each well of a new 96-well plate. To synchronize growth across conditions, resuspended cells were inoculated at medium-specific volumes: 3.7 µl (YPD), 4.0 µl (YPD10), 5.2 µl (YNB), and 5.0 µl (YEPG). For the slow-growing pTDH3 strain, an additional 2.5 µl of cell suspension was added in all media conditions. Plates were sealed with Breathe-Easy film prior to incubation. Absorbance at 600 nm (OD<sub>600</sub>), and sfGFP fluorescence, at 488 nm, was measured at 1 hour intervals in a plate reader (Biotek Synergy HT or Bionoy Enzo Absorbance 96), and OD<sub>600</sub> was monitored with the Platecurver script ([github.com/davinap/YTK-heterogeneity-platecurver](https://github.com/davinap/YTK-heterogeneity-platecurver)). Plate reader sfGFP intensity (PR) was scaled with the formula  $((PR \times 4.21 \times 10^{-5}) + 3.5)$  so it was comparable to the Log(sfGFP) intensity of the flow cytometry data ([github.com/davinap/YTK-heterogeneity-GFP](https://github.com/davinap/YTK-heterogeneity-GFP), [github.com/davinap/YTK-heterogeneity-flow](https://github.com/davinap/YTK-heterogeneity-flow)), to produce the plots in Figure 1. At the time points outlined in Tables S13 and S14 and Figures S16 and S17 in the Supplemental information online, 10 µl of each sample was washed two times in 200 µl of PBS, with centrifugation at 3000 g for 2 min in between (Eppendorf 5810). All samples were resuspended in 200 µl of 1 µg/ml propidium iodide in PBS (Merck P4170; Gibco 10010023). For each media condition, 3 compensation controls were used: the Control strain, the pCCW12-sfGFP strain, and Control strain heat-killed at 70 °C for 2 minutes. All were resuspended in 200 µl of PBS. Flow cytometry was conducted in the Attune Nxt (Invitrogen) using the following settings: FSC 100 V, SSC 340 V, BL1 400 V, YL2 480 V, FSC threshold of 10.0 × 1000, stop at 50,000 events. Flow cytometry was repeated for a total of three biological replicates, for pYTK strains, or five biological replicates, for the strains used to validate model predictions.

### Subpopulation analysis of strains

Flow cytometry data were saved as FCS files and analysed first in FlowJo™ (BD Life Sciences). FSC-H and FSC-A were plotted to gate for single cells in the Control strain population. The singlets gate was applied to each sample in the same media condition. Compensation was performed to account for spillover between the BL1-H and YL2-H channels with the following strains: Control PBS (BL1<sup>-</sup>/YL2<sup>-</sup>), Control heat-killed (BL1<sup>-</sup>/YL2<sup>+</sup>), pCCW12-sfGFP (BL1<sup>+</sup>/YL2<sup>-</sup>). The resulting compensation matrix was applied to all samples from the same media condition. Live cells were selected by gating for YL2<sup>-</sup> cells - propidium iodide stains the nuclei of dead cells and this can be detected as emission in the 620/15 nm range in the YL2-H channel. Cells with YL2-H intensity greater than  $2 \times 10^3$  were removed from further analysis. The full backgating process is illustrated in Figure S15. CSV files of the live population of cells and their BL1-H values - a proxy for sfGFP intensity - were further analysed in R to determine the number of subpopulations ([github.com/davinap/YTK-heterogeneity-flow](https://github.com/davinap/YTK-heterogeneity-flow)). Briefly, the distribution of sfGFP intensity was analysed using Hartigan's dip test ([github.com/mmaechler/diptest](https://github.com/mmaechler/diptest)), which generated a p-value for non-normal distributions. A value less than 0.05 indicated a multimodal distribution. Samples which passed the dip test were then analysed using the flexmix package [39], a finite mixture modelling package, to determine whether 2 or 3 subpopulations were present within a single sample. The framework of JEFworks Lab was followed ([jef.works/blog/2017/08/05/a-practical-introduction-to-finite-mixture-models/](http://jef.works/blog/2017/08/05/a-practical-introduction-to-finite-mixture-models/)). Across the three biological replicates, a consensus table was generated after manual curation to remove anomalous results which would otherwise interfere with the calculation of metrics like the mean of each subpopulation. Example analyses for the corresponding figures (Figures 1 and 2, Figures S1, S2, S11, and S12, and Tables S1, S2, S5, and S6) can be obtained at [github.com/davinap/YTK-heterogeneity-analysis](https://github.com/davinap/YTK-heterogeneity-analysis), [github.com/davinap/YTK-heterogeneity-GFP](https://github.com/davinap/YTK-heterogeneity-GFP), and [github.com/davinap/YTK-heterogeneity-model](https://github.com/davinap/YTK-heterogeneity-model). FCS files, plate layouts and analyses for all samples and biological replicates can be made available upon request.

## Analysis of growth arrested cells

### Flow cytometry

In a 96-well plate, 125  $\mu$ l of YPD was inoculated with  $\sim$ 1 mm of yeast glycerol stock of the corresponding strains in Table S8. This was repeated for a total of three biological replicates. A Breathe-Easy film (Sigma-Aldrich Z380059) was applied on top and the strains were incubated for 48 h at 30 °C with 700 rpm shaking (Infors Multitron with tray for microtiter plates). After cells reached saturation, the plate was spun down at 3000 g for 2 min (Eppendorf 5810). The supernatant was removed and cells were resuspended with 150  $\mu$ l of sterile PBS (Gibco 10010023). Cells were spun down and washed for a total of 3 washes. Cells were finally resuspended in 125  $\mu$ l of PBS. In a fresh 96-well plate, 125  $\mu$ l YPD was inoculated with 3.7  $\mu$ l of cell suspension. For each biological replicate, two rows were inoculated - one for treatment and one as an exponential phase control. A Breathe-Easy film (Sigma-Aldrich Z380059) was applied on top and the strains were incubated at 30 °C with 700 rpm shaking (Infors Multitron with tray for microtiter plates). OD<sub>600</sub> was monitored every hour in a Byonoy plate reader (Enzo, Absorbance 96) and the Platecurver script in R ([github.com/davinap/YTK-heterogeneity-platecurver](https://github.com/davinap/YTK-heterogeneity-platecurver)). At 6 h post-inoculation, one row per biological replicate was treated with 1.875  $\mu$ l of 200  $\mu$ g/ml -factor (The Francis Crick Institute Peptide Chemistry STP) for 2.5 h while the duplicate row was left to reach exponential phase. At 8.5 h, 15  $\mu$ l of each sample was added to 200  $\mu$ l PBS in a fresh plate. A lid was placed on the plate and it was centrifuged at 3000 g for 2 min (Eppendorf 5810). The supernatant was removed and cells were resuspended in 200  $\mu$ l of PBS. This wash step was repeated once more, before all samples were resuspended in 200  $\mu$ l of 1  $\mu$ g/ml propidium iodide in PBS. For each condition, two wells were used as compensation controls. These wells contained the Control strain (which had no fluorescence signal) or pCCW12-sfGFP (which was used as the sfGFP positive control); these cells were resuspended in 200  $\mu$ l of PBS only. Flow cytometry was conducted in the Attune Nxt (Invitrogen) using the following settings: FSC 100 V, SSC 340 V, BL1 400 V, YL2 480 V, FSC threshold of 10.0  $\times$ 1000, stop at 50,000 events.

### Data analysis

In FlowJo (BD Life Sciences), FSC-H and FSC-A were plotted to gate for single cells in the Control population. The singlets gate was applied to each sample in the same treatment condition. Then, compensation was performed to account for spillover between the BL1-H and YL2-H channels with the following strains: Control PBS (BL1<sup>-</sup>/YL2<sup>-</sup>), Control heat-killed (BL1<sup>-</sup>/YL2<sup>+</sup>), pCCW12-sfGFP (BL1<sup>+</sup>/YL2<sup>-</sup>). The resulting compensation matrix was applied to all samples for the same treatment. Live cells were selected by gating for YL2<sup>-</sup> cells - cells with YL2-H intensity greater than  $2 \times 10^3$  were removed from further analysis. The full backgating process is illustrated in Figure S15. Then the bisector tool was used to gate for low sfGFP (Pop1) and high sfGFP populations (Pop2) based on the location of the first trough in the Comp BL1-H distribution plot. Proportion of cells in Pop1 and Pop2 was recorded in FlowJo and Welch's t-test was performed in R ([github.com/davinap/YTK-heterogeneity-GA](https://github.com/davinap/YTK-heterogeneity-GA)) between treated and exponential phase groups across the three biological replicates.

## Cell cycle microscopy analysis

Promoter strains were grown as described in the growth arrest methods section for 48 h and subsequently resuspended in fresh YPD. After 4 h of growth, cells were treated with Hoechst 33342 at a final concentration of 12  $\mu$ g ml<sup>-1</sup> and incubated at 30 °C for an additional 2 h. Cells were then transferred onto microscopy slides (2.5  $\mu$ l per sample) and imaged using a Nikon Eclipse Ti-E inverted microscope equipped with a Hamamatsu ORCA-Flash4 sCMOS camera and a CoolLED pE-300 LED light source. Hoechst-stained DNA was visualised using the blue fluorescence channel (DAPI filter set; excitation 360 nm, emission 450 nm), while sfGFP fluorescence was detected using the green fluorescence channel (GFP/FITC filter set; excitation 480 nm, emission 525 nm). For each field of view, z-stacks were acquired at five focal planes spanning a total depth of 3  $\mu$ m. Bright-field images were collected for morphological reference. Images were processed and analysed using ImageJ/Fiji [64]. To avoid staining- or thresholding-related bias, cell cycle classification was performed manually based on a combination of Hoechst signal intensity and nuclear morphology, together with cell shape and budding status, allowing assignment of individual cells to G<sub>1</sub> or S/G<sub>2</sub>/M phases. An average of 50 cells were counted per condition (Table S9).

## Protein level prediction from scRNA-seq data

Single-cell RNAseq datasets were obtained from ref [41]. Data transformation was performed according to the publication. In brief, raw counts from wild-type diploid strains were obtained for growth on rich and minimal media (YPD and MMD datasets, respectively). The lognorm function from the Scuttle R package was used for read normalisation [63]. For each gene, promoter strength was computed as the total number of reads for scRNAseq datasets and weighted sums for sfGFP datasets. Promoter heterogeneity

was computed as coefficient of variations (CV), calculated as the standard deviation divided by the mean of the gene. Gaussian processes were used to predict sfGFP levels as a function of scRNAseq levels using the R package GauPro ([github.com/CollinErickson/GauPro](https://github.com/CollinErickson/GauPro)) on min-max normalised sfGFP and scRNAseq data. All scripts used to perform this analysis are available on GitHub ([github.com/KiyanShabestary/YTK-heterogeneity](https://github.com/KiyanShabestary/YTK-heterogeneity)).

### Sample preparation for Raman spectroscopy

Strains Hom1, Het1, Hom2, Het2 and Control (Table S8) were streaked on YNB plates, from glycerol stocks, and grown at 30 °C to obtain single colonies. A single colony was used to inoculate a 96-well plate containing 125 µl YNB media. The plate was sealed with Breathe-Easy film and incubated at 30 °C for 24 h with shaking at 700 rpm (Infors Multitron). The next day, in a fresh 96-well plate, 125 µl of YNB was inoculated with overnight culture with the following volumes: 0.4 µl of Hom1, 5.2 µl of Het2, 0.4 µl of Hom2, 0.4 µl of Het2 and 0.2 µl of Control. For each growth phase, there were 4 biological replicates. The plate was sealed with Breathe-Easy film and OD<sub>600</sub> was monitored in a Byonoy plate reader (Enzo Absorbance 96), with the plate replaced in the shaking incubator in between readings. Exponential phase samples were taken at 22.75 h whereas stationary phase samples were taken at 40 h. 90 µl of each sample was added to 10 µl of 37 % formaldehyde (Sigma-Aldrich F8775) in a 1.5 ml tube. Tubes were rotated end-on-end in a rotator (StarLabs RM-Multi 1) at room temperature for 14 min with the following settings: 15 rpm orbital, 15° reciprocal, 5° vibro. Cells were spun down at 6000 g for 1 min. Formaldehyde was removed and 100 µl of ice-cold PBS was added before gentle vortexing to wash cells. Cells were spun down at 6000 g for 1 min, the supernatant was removed and the wash process was repeated for a total of 3 washes. Cells were resuspended in 100 µl of ice-cold PBS and stored at 4 °C until Raman spectroscopy.

### Single-cell lycopene measurements using Raman spectroscopy

Cells stored in cold PBS were washed with ddH<sub>2</sub>O for three times and dropped onto a Raman slide with aluminium coating for air drying before Raman spectra of single cells of Control, Hom1, Het1, Hom2 and Het2 were acquired by a HORIBA Raman microscopic spectrometer (HR Evolution). A 532 nm laser at 4 mW power was focused onto a cell through a 25 % neutral density (ND) filter with a 50× objective (NA=0.5) with a 100 hole to make sure a single laser spot could cover the size of one single cell. The exposure time was set to 100 ms per acquisition due to the strong resonance effect of the lycopene signal. The Raman grating used for single-cell measurements was 300 gr/mm. For each strain, >200 cells were measured across four biological replicates. In total, 8,950 single-cell Raman spectra were obtained. Preprocessing for the raw Raman spectra included quality control for eliminating abnormally/burnt high-intensity spectra, cosmic ray correction, baseline fitting and subtraction for autofluorescence removal. Quantification of lycopene was done by calculating the sum of the integrated areas of three lycopene bands centred at 1002, 1154 and 1511 cm<sup>-1</sup>. The threshold of determining whether the cells contain lycopene was set to the lycopene intensity in the Control strain. Single-cell lycopene intensities were used to visualise lycopene distribution among strains; averaged lycopene intensities from all single cells in one strain were used for comparison between strains.

### Lycopene quantification using HPLC and FlowRACS

Strains Hom1, Het1, Hom2, Het2 and Control (Table S8) were streaked on YNB plates, from glycerol stocks, and grown at 30 °C to obtain single colonies. A single colony per strain and replicate was used to inoculate as precultures in 3 ml tubes and subsequently transferred to 15 ml flasks for 40 h. After incubation, 10 ml of culture was harvested by centrifugation at 4,000 rpm for 5 min. Cell pellets were resuspended in 1 ml ice-cold acetone and subjected to lycopene extraction using a tissue homogeniser (six cycles of 1 min at 8,000 rpm with intermittent cooling). Following extraction, lycopene was quantified using a Vanquish KITT HPLC system (Thermo Fisher Scientific) equipped with a YMC Carotenoid C30 column (150 mm × 4.6 mm, 3 µm). The mobile phase consisted of solvent A (methanol), solvent B (ethyl acetate), and solvent C (tert-butyl methyl ether) in a ratio of 1200:400:1000 (v/v). Quantification was performed against a pure lycopene standard dilution curve. Peak analysis and integration were carried out using Chromeleon Software (Version 7, Thermo Fisher Scientific). The experiment was performed using 3 biological replicates.

For FlowRACS analysis, 1 µl of shakeflask cultures were diluted in 5 ml FlowRACS buffer and subsequently loaded onto microfluidic chips that enable cell focussing under the Raman laser (532 nm) as described by Wang and colleagues [66,67]. Raman spectra were then acquired continuously at a flow rate of 40 µl/min. The laser power was set to 60 mW at an acquisition time of 0.5 s per cell. Files containing single-cell Raman spectra were imported into RStudio and compared to the spectrum of a pure Lycopene spectrum. For Lycopene-specific peaks, the area under the curve was calculated and the sum of all peak areas was compared between different samples.