Inheritance of expression reveals patterns of quantitative trait inheritance in yeast

Donya Khoury, Noa Aharon-Hefetz, Bar Cohen, Yitzhak Pilpel, Orna Dahan

Department of Molecular Genetics, the Weizmann Institute of Science.

Abstract

Understanding how quantitative traits are inherited remains a central question in genetics. Although gene expression levels and their cell-to-cell variation (i.e. "noise") are well studied, their inheritance in sexually produced offspring is still underexplored. Here, we map the genetic architecture of expression level and expression noise using a large-scale half-diallel design in *Saccharomyces cerevisiae*. We crossed 99 MATa and 45 MATα natural strain isolates engineered to express GFP or RFP, generating >2,800 diploid combinations carrying both markers. Single-cell flow cytometry in parents and offspring enabled precise estimates of mean expression and expression noise of all parents and offspring. We find that mean expression is heritable and shaped by both parental contributions and mating-type–specific effects. Beyond classical additivity, we identify two modes of inheritance: **parental inheritance**, in which a parent influences the expression of the allele it contributes, and **non-parental inheritance**, in which a parent affects expression of the allele inherited from its mate. Notably, we demonstrate that gene-expression noise is a heritable trait following sexual mating, a property previously emphasized mainly in asexual contexts. Machine-learning models trained on parental transcriptomes and strain-level metadata show that noise is genetically encoded and inherited. We revealed a set of genes whose expression level in parents predicts with high accuracy the noise level in their offspring, revealing a first indication for a heritable mechanism that governs stochastic cell-to-cell variation in gene expression. Together, these results establish that both expression level and noise are heritable, genetically encoded traits whose inheritance can follow complex, non-additive patterns. Our framework provides a scalable route to dissect regulatory variation in diverse genetic backgrounds and clarifies how expression traits, and their variability, are transmitted across generations.

Introduction

The inheritance of traits across generations of sexually reproducing organisms is a cornerstone of genetics, shaping our understanding of evolution, and the subtle forces that govern life's diversity. While Mendel's laws provide a framework for simple traits, the inheritance of quantitative traits - shaped by multiple genes, complex regulation, and environmental influences - is far more complex. Classic examples such as height or metabolic rate exhibit continuous variation and are often unpredictable from parental genotypes alone (Falconer, 1996). Gene expression is one of the most accessible and

informative quantitative traits, offering a powerful model for studying the genetic basis of complex phenotypes. It is influenced by both cis- and trans-acting genetic variants (Rockman & Kruglyak, 2006), and can be measured with high precision (Albert & Kruglyak, 2015). Studying gene expression as a quantitative trait enables the dissection of regulatory architecture and provides insight into how genetic variation translates into functional differences at the molecular level (Lappalainen et al., 2013). Beyond mean expression, gene expression noise—the cell-to-cell variability in expression among isogenic cells in the same environment— has emerged as a crucial yet understudied trait. Early works (Bar-Even et al., 2006a; Newman et al., 2006) linked noise to promoter architecture and regulatory mechanisms, while more recent studies (You et al., 2019) showed that noise-buffering capacity is genetically encoded and subject to selection. Although (Ansel et al., 2008) provided early evidence that gene expression noise can be inherited following sexual reproduction, their scope was limited to a small number of parental crosses and haploid recombinant offspring, leaving open the question of how noise is inherited in diploids. To investigate the genetic basis of both expression levels and noise, we use *Saccharomyces cerevisiae* (baker's yeast) as a model organism. Budding yeast offers several key advantages: a fully sequenced and well-annotated genome (Goffeau et al., 1996; Parapouli et al., 2020), rapid growth allowing high-throughput culture, allowing for high-throughput experimentation, and extensive natural diversity (Peter et al., 2018), that has been leveraged to map genotype–expression relationships (Brem et al., 2002). A unique feature of *S. cerevisiae* is its ability to exist and proliferate as both haploids and diploids, proliferating vegetatively through mitotic division in both haploid and diploid states. Sexual reproduction involves fusion of haploid cells of opposite mating types (MATa and MATα) via a pheromone-mediated process (Herskowitz, 1988). During mating, the genetic material from two parents combines, resulting in a diploid organism that inherits traits from both mating types (Sherman, 1991). Because haploids are fully functional vegetative cells, their phenotypes can be directly compared to diploid offspring—something impossible in most other eukaryotes. This property provides a unique opportunity to study how gamete phenotypes relate to zygote traits. The inheritance of gene expression levels in yeast has been widely explored. Foundational work (Brem et al., 2002) identified cis- and trans-acting variants influencing gene expression in diploid progeny from a cross between a laboratory and a wild strain, and later studies (Smith & Kruglyak, 2008) examined expression heritability and gene-environment interactions, and allele-specific expression (Khan et al., 2012; Ronald & Akey, 2007) using diploid hybrids to distinguish cis from trans effects. These studies, however, typically involved a limited number of parental strains and used bulk transcriptomics, limiting the resolution at which expression variation—and especially stochasticity—could be observed. More recently, (Tsouris et al., 2024) used a half diallel cross with diverse isolates to quantify global expression inheritance, revealing substantial non-additive components.

Our study builds on and extends this framework. We systematically crossed over 100 genetically diverse haploid parents to generate more than 2,800 genetically distinct diploid offspring using the half

diallel approach, enabling analysis of inheritance across a broad genetic landscape. We focus not only on mean expression but also on gene expression noise—a trait rarely examined in sexual inheritance. To do so, we engineered each haploid mating type to express a distinct fluorescent protein (GFP or RFP) from an identical, defined locus. This dual-marker system allows us to trace each reporter independently and disentangle parental versus trans-acting influences in offspring. This design allows us to capture novel inheritance patterns not accessible to traditional eQTL or hybrid-based studies. Specifically, we define two distinct modes of inheritance: *parental inheritance*, where a parent influences the expression of the marker it contributed, and *non-parental inheritance*, where a parent influences the other parent's marker. Our data reveal novel inheritance patterns for both expression level and noise. Integrating single-cell phenotyping with transcriptomics and machine learning, we identify genetic features predictive of these traits and demonstrate that gene expression noise is a heritable, genetically controlled trait transmitted through sexual mating. This integrative approach offers a novel framework for dissecting the inheritance of complex traits with implications for evolutionary biology, synthetic biology, and biotechnology.

Results

Natural Variation in Fluorescent Gene Expression Across Haploid Strains

To investigate the inheritance of quantitative traits, we focused on two phenotypic categories: gene expression and expression noise. These traits were chosen for their measurable, continuous distributions and their relevance to understanding the heritability of molecular phenotypes in *Saccharomyces cerevisiae*.

To explore the inheritance of these traits, we utilized a system previously developed in our laboratory (Strauss et al., 2024). Briefly, we employed a collection of *Saccharomyces cerevisiae* natural isolates (Peter et al.) to generate 99 MATa haploid strains, which were genomically engineered to express a heterologous GFP protein. Expression was driven by the strong constitutive *TEF1* promoter, with the construct integrated at the *HO* locus. In parallel, our strain collection included 45 Matα haploid strains in which the heterologous fluorescent protein RFP, also under control of the *TEF1* promoter, was integrated at the same genomic location (*HO* locus). The full list of strains is provided in Supplementary Table S1. The strains were derived from diverse geographic regions and ecological niches, encompassing a wide range of fitness levels and genetic diversity, with pairwise genetic distances ranging from 0 to 1.4 SNPs per kilobase (for additional details, see Strauss et al., 2024).

We began by measuring fluorescent level at a single cell resolution in all 144 haploid strains in 3 independent repetitions. A representative fluorescence distribution for selected haploid strains (and a randomly chosen diploid offspring discussed below) are presented in Supplementary Figure S1A–B. Reassuringly, we observed high correlations between the florescent levels measured at the different replicates (Figure S1C-D). Figure 1 shows the distribution of fluorescence intensities across 99 MATa

strains (panel A) and 45 MATα strains (panel B). Each dot represents a single cell measurement (~25,000 cells per strain), and the boxplots summarize the central tendency and variability within each strain. Colored dots indicate the mean fluorescence of each biological replicate. To account for potential effects of cell size on fluorescence measurements, we normalized fluorescence intensity by dividing it by the forward scatter area (FSC-A), a proxy for cell size. We observed a strong positive correlation between raw and normalized fluorescence levels ($r = 0.78$, $p < 10^{-300}$ for MATa; $r = 0.90$, $p < 10^{-300}$ for MATα; Supplementary Figure S1G–H), indicating that cell size does not account for the observed expression differences between strains.

The results revealed, a wide variation in fluorescence intensity among strains, spanning approximately 3-fold in MATa and 18-fold in MATα. To assess the statistical significance of expression differences between all parental strains, we performed a rank-sum test. This analysis revealed that the majority of pairwise comparisons between strains are statistically significant (Figure S2), indicating widespread expression variation within each mating type.

Our strain collection included 32 pairs in which both MATa and MATα haploids were derived from the same diploid natural isolate. Correlation analysis of GFP and RFP expressions across these strain pairs revealed no significant relationship. Since all haploid strains originated from sporulated diploid isolates, the absence of correlation may be partially explained by the segregation of heterozygous SNPs during meiosis.

To examine this possibility, we stratified the strains based on the level of heterozygosity in the original diploid isolates, as determined by (Peter et al., 2018)Notably, haploids derived from homozygous diploid backgrounds exhibited a modest yet significant positive correlation in GFP and RFP expression levels ($r = 0.34$, $p = 0.1$). In contrast, no correlation was observed among haploids derived from heterozygous diploids (Supplementary Figure S3A-C). Together, these findings suggest that differences in genetic background contribute to the observed variability in gene expression levels, as measured by the heterologous GFP and RFP reporters.

Analysis of Offspring Expression Profiles Reveal Complex Inheritance

Given that our analysis of MATa and MATα haploid strains revealed differences in the mean expression levels of the fluorescent proteins, this system offers a powerful platform for studying the inheritance of gene expression. To this end, we systematically crossed each MATa strain with each MATα strain to generate diploid strains harboring both GFP and RFP constructs on homologous chromosomes. In total, we generated a comprehensive library of 3,002 diploid strains, representing all successful mating combinations between the parental haploids. This dataset enabled a detailed investigation into the patterns and principles governing the inheritance of gene expression levels in *S. cerevisiae*.

We then systematically measured by FACS the expression levels of both fluorescent proteins in all resulting diploid offspring, organized in 96-well plates (~35K cells from each offspring strain), and calculated the geometric mean fluorescence across all measured cells, using it as the representative

expression level of each strain. To assess data reproducibility, fluorescence measurements were collected from two independent technical replicates for a subset of randomly chosen 800 strains. A strong correlation was observed between replicates for both GFP and RFP expression ($r = 0.82$, $p = 1.15 \times 10^{-198}$ for GFP; $r = 0.90$, $p = 1.6 \times 10^{-295}$ for RFP; Supplementary Figure S3D-E), confirming the robustness of the fluorescence measurements used in downstream analyses. In addition, to rule out batch or geographical effects, we plotted a heatmap of GFP and RFP expression levels organized by plate and well position. This analysis revealed no evidence of batch effects or spatial biases within or across plates (Supplementary Figure S3F–G).

To enable direct comparison between GFP and RFP fluorescence levels in the offspring, all measurements were normalized to the median fluorescence value of each experiment. This normalization accounted for differences in detection sensitivity between GFP and RFP channels during FACS analysis and ensured comparability across experiments. Normalized values were used for all subsequent analyses and visualizations of offspring expression.

Analysis of fluorescence distributions revealed that the majority of offspring spanned approximately a 2.5 -fold range in both GFP and RFP expression. However, a subset of offspring exhibited an additional two-fold increase in RFP levels (Figure 1C). Notably, offspring derived from the MATα strains AKP and BBI consistently showed high RFP expression, highlighted in yellow and blue, respectively. These strains are among the top RFP producers among the parents. In contrast, the two highest GFP-expressing MATa parents, BHA and CFR, did not produce offspring with similarly elevated GFP levels (highlighted in red and green, respectively) (Figure 1C). These findings indicate that, similar to the parental strains, diploid offspring display wide variability in fluorescent gene expression, while the similarity in expression across AKP and BBI derived offspring suggests a cis-regulatory effect from the MATα parent, the absence of additional similar cases especially among the MATa parents and indicates more complex regulatory influences. To systematically assess these patterns across the entire panel, we next conducted a correlation analysis between parental and offspring expression levels. As a first step, we assessed how the expression levels of the heterologous proteins in the parental strains relate to those in their offspring. For each parental pair, we computed the average fluorescence values of GFP (from the MATa parent) and RFP (from the MATα parent) and compared them to three corresponding measurements in the offspring: normalized GFP fluorescence, normalized RFP fluorescence, and the average of both fluorescent signals. This analysis allowed us to evaluate how parental expression levels influence offspring expression, and to test whether inheritance follows a symmetric average, or if expression is more strongly shaped by the parent contributing the corresponding fluorescent allele. To ensure robustness, offspring of the MATα strains AKP and BBI, identified as outliers due to their unusually high RFP expression, were excluded from this analysis (the full analysis including these outliers is provided in Supplementary Figure S4A-C). After outlier removal, we observed a moderate but statistically significant correlation between parental average fluorescence and the offspring fluorescence ($r = 0.11$, $p = 1.21 \times 10^{-7}$ for GFP; $r = 0.11$, $p = 3.46 \times 10^{-8}$ for RFP, Figure 1D-E). The

correlation was slightly stronger when considering the offspring average of both fluorescent markers ($r = 0.15$, $p = 1.21 \times 10^{-13}$; Figure 1F). Additional correlation analyses using alternative parental metrics including minimum, maximum, and other combinations of parental fluorescence values are provided in Supplementary Figure S4E–F., along with results from analyses that include the previously excluded outliers. These results suggest that the inheritance of fluorescent gene expression in diploid offspring does not follow a simple additive model—offspring expression levels cannot be accurately predicted by averaging the expression values of the two haploid parents. Moreover, we find that offspring expression is not well explained by the minimum or maximum parental expression either. This observation led us to ask whether distinct modes of inheritance—beyond straightforward additive contributions—could be identified among the offspring. This is consistent with recent findings by (Tsouris et al., 2024), who showed that approximately one-third of the genome in *Saccharomyces cerevisiae* do not follow additive inheritance patterns but instead exhibit substantial non-additive genetic components in their expression profiles.

Figure 1

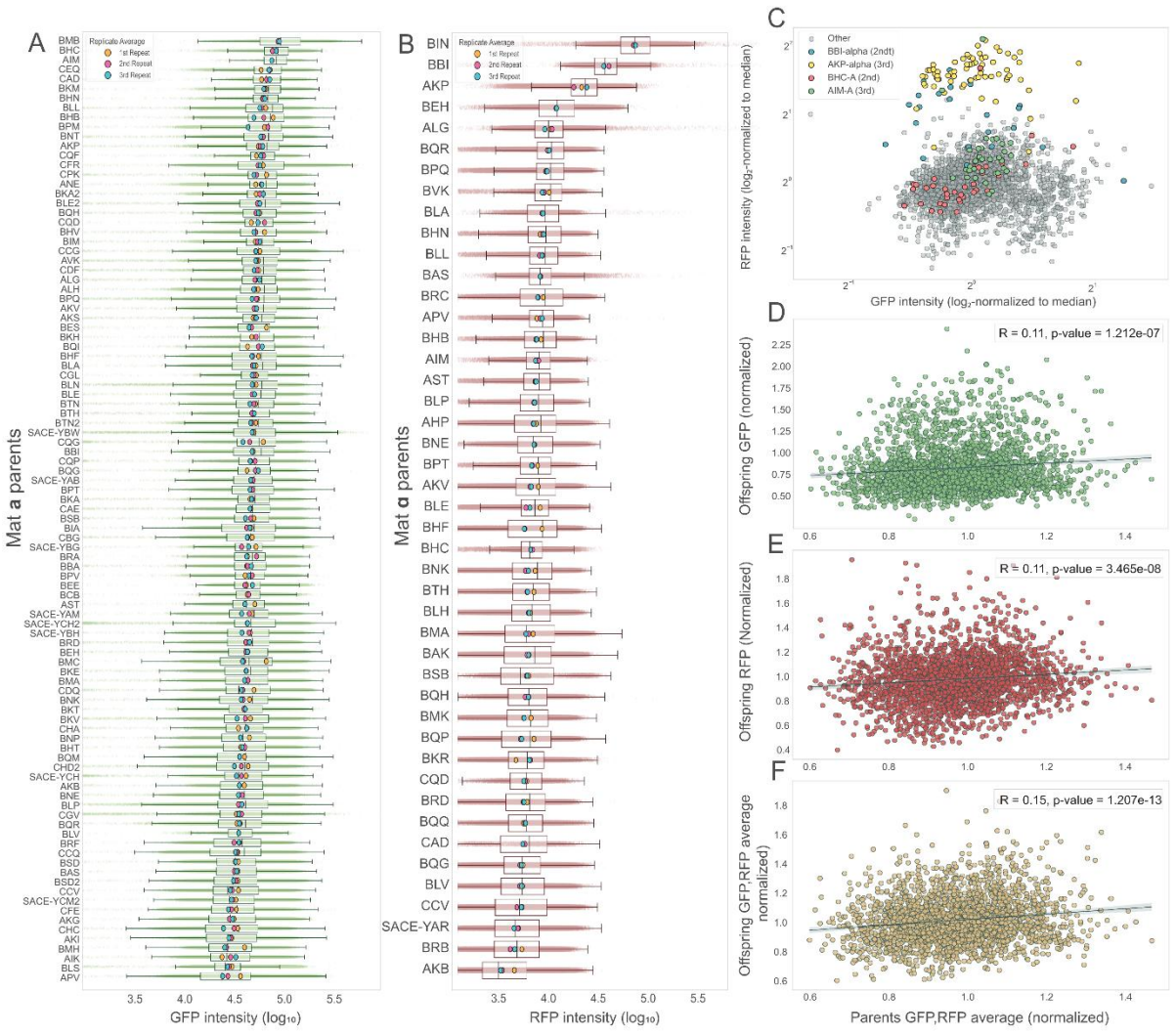

Figure 1 | Fluorescent protein expression levels varies across strains, and offspring expression moderately correlates with parental averages. (A) GFP fluorescence intensity (log_{10}-transformed, Y axis) measured in 99 MATa strains (X axis).

(B) RFP fluorescence intensity (log_{10}-transformed) measured in 45 MATα strains. In both A and B sub plots, each point represents a single cell, with ~75,000 cells measured per strain. Boxplots summarize the distribution of fluorescence values for each strain, and colored dots represent the mean fluorescence of three biological replicates (replicate 1: yellow, replicate 2: pink, replicate 3: cyan). Strains are ordered by decreasing mean fluorescence. (C) GFP and RFP expression levels (log_{10}-transformed) of 3,002 diploid offspring, normalized to the median fluorescence of all offspring for each marker. Each point represents a single offspring strain. Offspring derived from MATα strains AKP and BBI, characterized by elevated RFP expression in the parental strain, are highlighted in yellow and blue, respectively. Their corresponding MATa type counterparts with the highest GFP expression in the parental strain, BHA and CFR, are highlighted in red and green, respectively. (D–F) The average fluorescence values of GFP (from the MATa parent) and RFP (from the MATα parent), plotted against: Offspring GFP (D), RFP (E), and the average of GFP and RFP expression (F). Outliers (AKP- and BBI-derived offspring) were excluded from the correlation analysis. Moderate but statistically significant correlations were observed for all measures ($r = 0.11$, $p = 1.21 \times 10^{-7}$ for GFP; $r = 0.11$, $p = 3.46 \times 10^{-8}$ for RFP; $r = 0.15$, $p = 1.21 \times 10^{-13}$ for average fluorescence).

Identification of Parental and Non-Parental Inheritance Patterns

The observed differences in expression levels between strains with different genetic backgrounds, despite carrying the same heterologous protein construct (Figure 1A–B), together with the positive yet moderate correlation between the expression levels of the fluorescent genes in parents and their corresponding offspring (Figure 1D–F), suggest a complex genetic architecture underlying gene expression inheritance. To further explore the relations between the expression of the fluorescent genes in the offsprings and their respective parents we generated heatmaps of GFP and RFP expression levels across all offspring, organized by half-sibling group. This analyses revealed a distinct pattern of inheritance that we term *parental* and *non-parental* inheritance. *Parental inheritance* refers to cases where a parent influences the expression level of the fluorescent marker it contributed to the offspring—for example, a MATα parent consistently producing offspring with elevated RFP expression. *Non-parental inheritance*, by contrast, describes instances where a parent affects the expression of the marker contributed by the other parent. For example, certain MATa strains—donating the GFP marker—produced offspring that consistently exhibited high RFP expression. The conceptual distinction between these two modes of inheritance is illustrated in Figure 2C.

(Figure 2A–B) in these heatmaps, MATa parents are arranged on the vertical axis and MATα parents on the horizontal axis, enabling visualization of offspring that share a common parent. Since each diploid inherits one set of chromosomes—and one fluorescent marker gene—from each haploid parent, this layout allows us to distinguish whether a parent's influence is on its own marker (*parental inheritance*) or on the partner's marker (*non-parental inheritance*). For example, parental inheritance was observed in the case of MATα parent AKP, which consistently produced offspring with elevated RFP levels—the marker it contributed. In contrast, non-parental inheritance was evident in MATa strains such as AIM and AKI, which contribute GFP but gave rise to offspring with consistently high RFP expression, indicating that these parents influence the partner's marker. This heatmap-based approach revealed consistent expression trends among the offspring of specific strains. Strains that repeatedly gave rise to offspring with similarly high or low expression—whether for the marker they contributed or the one from their mate—are highlighted with dashed black rectangles in Figure 2A–B.

To systematically identify and quantify cases of *parental* and *non-parental inheritance*, we compared each parent's own fluorescence intensity to the mean expression level of the corresponding marker in

all of its offspring. Figure 2E presents RFP expression of the offspring that share the same MATα parent but differ in their MATa parent. Offspring are arranged along the x-axis and grouped by their MATa parent, with expression levels displayed as boxplots representing half-sibling groups. This analysis revealed clear evidence of *parental inheritance*, where the parent influences the expression of the marker it contributes to the offspring—particularly in the case of strains such as AKP and BBI. These MATα parents, which were among the highest RFP expressers as haploids (Figure 1B), consistently produced offspring with similarly elevated RFP levels, illustrating a direct influence on the expression of their own marker (Figure 2E). Similarly, Figure 2D shows GFP expression in offspring that share the same MATa parent but differ in their MATα parent. Certain MATa strains such as BES, SACE-YBG, and BCB consistently produced offspring with elevated GFP expression, indicating a pattern of *parental inheritance*. However, these strains displayed only moderate GFP levels in the haploid state (Figure 1A). This observation led us to investigate whether cis- or trans-acting regulatory factors could underlie the observed variation in fluorescent marker expression among offspring. To assess potential cis-regulatory contributions, we sequenced the genomic regions surrounding the integrated fluorescent gene—including the promoters and flanking sequences—in strains selected for their differing fluorescence levels. This analysis revealed no sequence variation in the promoter or integration site regions, suggesting that cis-acting genetic differences are unlikely to explain the inheritance patterns observed in the diploids. Further inspection of the distributions revealed cases of *non-parental inheritance*. For example, MATα strains BHB and BQP were consistently associated with offspring exhibiting low and high GFP expression, respectively—despite these parents not contributing the GFP marker. These effects are highlighted in orange and blue, respectively (Figure 2D). Similarly, MATa strains AKI and BMA influenced RFP expression levels in their offspring, despite not contributing the RFP marker; these are likewise highlighted in orange and blue in Figure 2E. Interestingly, three of these four parental strains—BQP, AKI, and BMA—exhibited only moderate expression of their own fluorescent marker as haploids, yet, they consistently drove extreme expression levels of the other fluorophore in their offspring.

This pattern suggests that *non-parental inheritance* may be mediated by trans-acting regulatory factors that modulate gene expression across the diploid genome—possibly through transcription factors or cofactors acting on alleles inherited from the mating partner. One hypothesis was that such trans-acting factors might be linked to TEF2 or TDH3, the endogenous promoters used to drive RFP and GFP expression, respectively. If so, strains with naturally high expression of these genes could produce elevated levels of transcriptional machinery that enhance expression of the matching construct—even when it is inherited from the opposite parent. To test this, we examined transcriptomic data from (Caudal et al., 2024) and found no consistent association between parental TEF2 or TDH3 expression and the strength of parental or non-parental inheritance effects. These results suggest that general overexpression of these endogenous genes is not a primary driver of the observed trans-regulatory inheritance, pointing instead to other, trans-acting factors as likely contributors.

Given that fluorescent gene expression exhibits clear non-additive inheritance patterns in our data, this observation aligns with (Tsouris et al., 2024), which showed that non-additive variation in gene expression in *S. cerevisiae* is predominantly driven by trans-regulatory mechanisms targeting core, highly connected cellular processes. While cis-trans compensatory interactions were also observed, trans-only effects were both more frequent and had stronger phenotypic impact. Our findings support and extend this model, reinforcing the view that trans-regulatory architecture plays a central role in mediating parental and non-parental inheritance and shaping non-additive gene expression in diploid yeast.

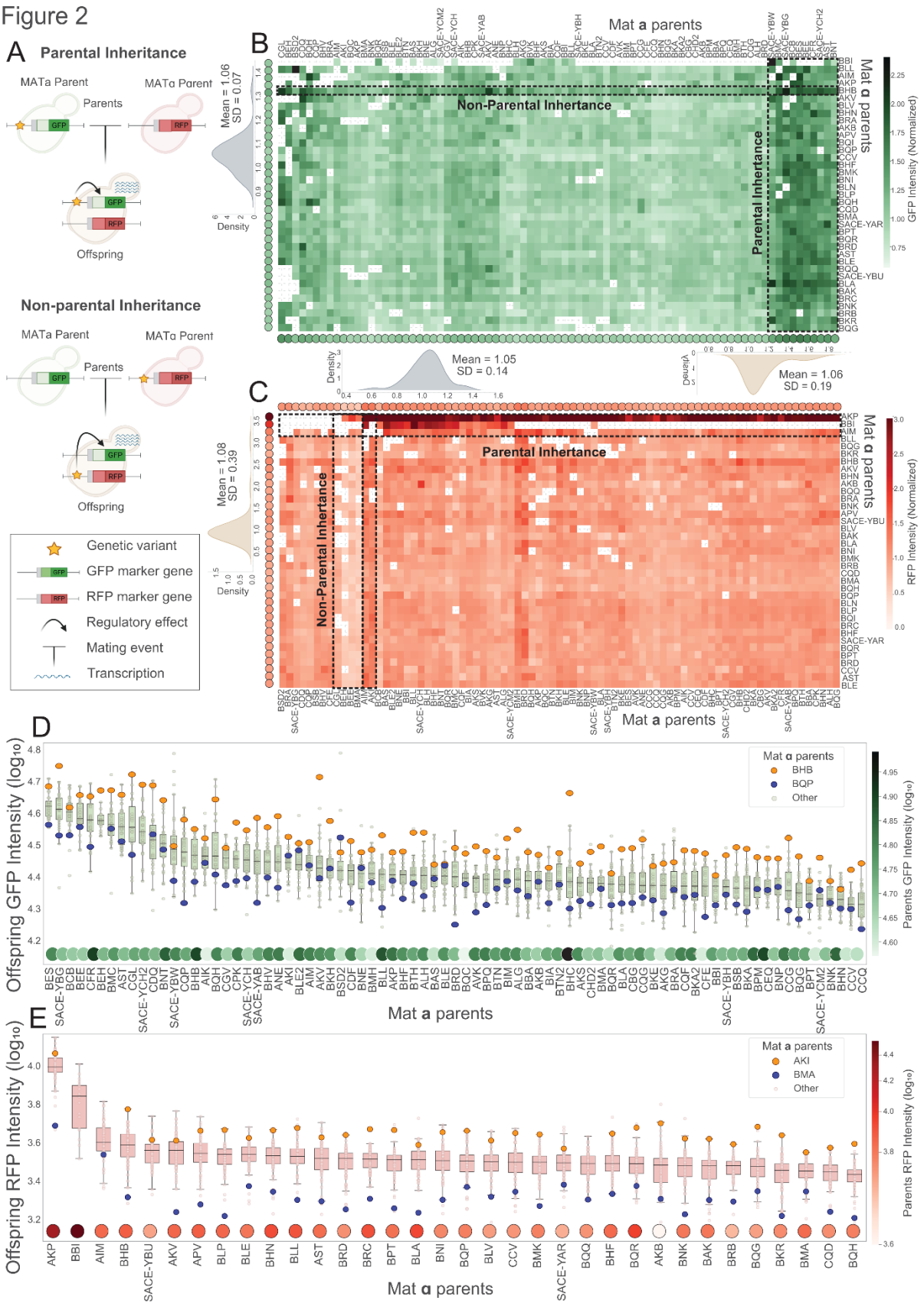

Figure 2 | Offspring mean expression levels reveal both parental and non-parental inheritance patterns. (A–B) Heatmaps of normalized GFP (A) and RFP (B) expression levels across all diploid offspring. Offspring are organized by half-sibling groups: MATa parents are arranged along the vertical axis, and MATα parents along the horizontal axis. Dashed black rectangles highlight groups of half-siblings consistently exhibiting high or low expression patterns, corresponding to cases of

parental or non-parental inheritance. (C) Conceptual illustration of parental and non-parental inheritance modes. In parental inheritance, a parent influences the expression level of the fluorescent coding gene it contributed to the offspring (e.g., an Alpha parent affecting RFP expression). In non-parental inheritance, a parent influences the expression of the fluorescent coding gene contributed by the opposite mating type (e.g., an A parent influencing RFP expression). fluorescent expression levels (D) GFP and (E) RFP in offspring of individual parents. Offspring are grouped by their MATa parent (D) and MATα parent (E). Each point represents a single offspring strain; boxplots summarize the distributions. Colored dots beneath each boxplot indicate the parent's own mean fluorescence intensity. In panel (D), MATα parents BHB and BQP (highlighted in orange and blue, respectively) were associated with consistently high and low GFP expression in their offspring, respectively, despite not contributing the GFP marker. In panel (E), MATa parents AKI and BMA (highlighted in orange and blue, respectively) influenced offspring RFP expression despite not contributing the RFP marker.

Parental Effect and Predictive Features of Fluorescent Expression in Offspring

To identify the most predictive features of GFP and RFP expression levels in the offspring, we trained a regression model using the DoubleLearningCatBoost approach—a two-phase CatBoostRegressor training strategy. In the first phase, the model was trained on a subset of the data (80%), and then refined in a second phase using a warm-started approach, which allowed the model to retain information from earlier training. We then evaluated its ability to predict expression in the held-out 20% of offspring not seen during training. To interpret the model's predictions, we applied SHAP (SHapley Additive exPlanations), which quantifies the contribution of each feature to the model's output in a biologically interpretable way. The features used to train the model are listed in the Materials and Methods section. Model performance was evaluated using both Pearson correlation and the coefficient of determination (R^2) (Predicted vs. actual scatter plots are shown in Supplementary Figure 5SA-B). When predicting mean expression levels, the model achieved an R^2 of 0.75 and a Pearson correlation of 0.87 for GFP, and an R^2 of 0.90 and a Pearson correlation of 0.95 for RFP. These results indicate strong predictive performance for both fluorescence markers. A pronounced parental contribution was evident in the features selected by the model. For GFP expression, 15 of the 16 top-ranked predictor genes originated from the MATa parent, which is also the contributor of the GFP marker (Figure 3A). Similarly, for RFP expression, 10 of the 15 most important genes were contributed by the MATα parent, which carries the RFP marker (Figure 3B). Interestingly, cell size emerged as the single most important predictor of both GFP and RFP expressions, highlighting a shared biophysical determinant of fluorescence across markers. Moreover, SHAP analysis revealed a positive correlation between cell size and expression levels: larger cells were consistently associated with higher predicted fluorescence for both markers. Among gene-level features, YHL015W.A, a gene of currently uncharacterized function, was the strongest predictor for GFP expression. In contrast, the strongest predictor for RFP expression was YMR020W (FMS1), a polyamine oxidase involved in converting spermine to spermidine. a polyamine oxidase involved in converting spermine to spermidine, a key step in the hypusination of the translation factor eIF5A, a modification required for cellular proliferation and translation (Joets et al., 1996) (Figure 3).

Figure 3

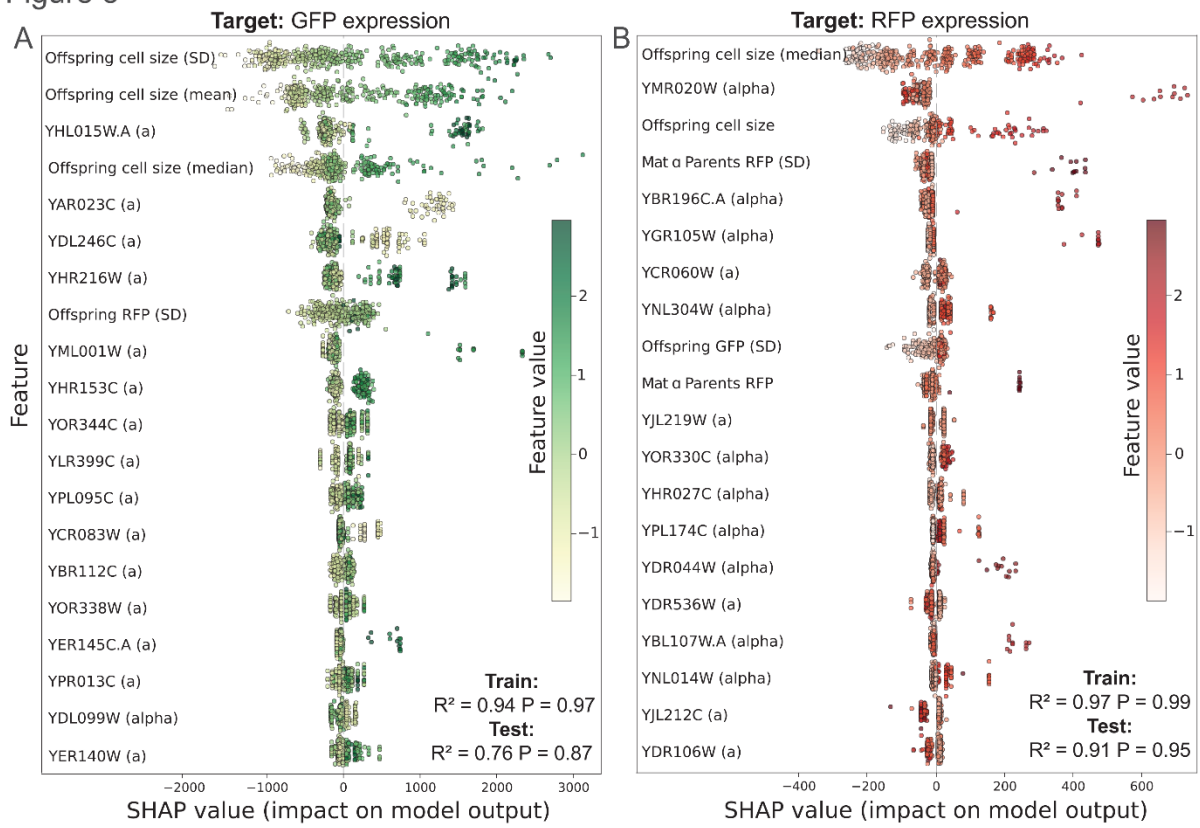

Figure 3 | SHAP summary plots identify predictors of offspring fluorescent expression and reveal parental effect. (A) SHAP values for the model predicting offspring GFP mean expression. Each dot represents a sample, and its horizontal position indicates the impact of that feature on the model's output. Features are ranked by overall importance (mean absolute SHAP value). Green color indicates the scaled value of the input feature (low to high). The model achieved $R^2 = 0.76$ and Pearson correlation = 0.87 on the test set. (B) SHAP values for the model predicting offspring RFP mean expression, following the same structure as (A). The model achieved $R^2 = 0.91$ and Pearson correlation = 0.95 on the test set. For both targets, offspring cell size (forward scatter) was among the top predictive features and showed a positive correlation with expression levels, indicating that larger cells tended to exhibit higher fluorescence. A strong parental effect is observed: in the GFP model (A), 15 of the top 16 predictive genes originated from the MATa parent; in the RFP model (B), 10 of the top 15 genes came from the MATα type parent, matching the parental origin of the respective markers.

Noise in Gene Expression Reflects Heritable Properties Following Sexual Mating

In addition to mean expression levels, our system enabled us to examine also the inheritance mode of gene expression variability between isogenic cells i.e noise. Similar to our analysis of mean expression levels we used our dual-reporter yeasts to investigate how gene expression noise is distributed across natural isolates and how this trait is inherited in diploid offspring.

To quantify expression variability in our system, we quantified expression noise for both GFP and RFP using the squared coefficient of variation (σ^2/μ^2). For each strain, we calculated the mean and standard deviation of fluorescence across the population of cells measured by FACS (on average, ~25,000 cells per strain) and used these values to compute noise. Noise was measured separately for MATa strains (expressing GFP) and MATα strains (expressing RFP). To characterize the relationship between expression level and noise, we plotted noise against mean fluorescence and fitted linear models after excluding outliers, following the approach described by Bar-Even et al., 2006 (see Materials and Methods for details). For both MATa (GFP) and MATα (RFP) parental strains, we observed a strong

negative correlation between mean expression level and extent of noise (slopes of –0.93 and –0.82, respectively, Figure 4A-B). This observation is consistent with previous reports showing that protein abundance inversely correlates with expression noise (Bar-Even et al., 2006).

Notably, several parental strains deviated from this overall trend. These strains were identified as visual outliers based on their clear deviation from the global pattern (see Materials and Methods for details). MATa strains SACE-YBW, CFR, and CGV (highlighted in navy, pink, and orange, respectively) exhibited elevated noise relative to their GFP expression levels. In addition, MATα strains AKP and BBI (highlighted in yellow and light blue) showed similar deviations in RFP noise (Figure 4A–B).

Offspring exhibited a similar trend to the parents, showing a negative correlation between expression noise and mean fluorescence for both GFP and RFP (Figure 4C-D). Notably, offspring derived from the previously identified high-noise parental strains—MATa strains SACE-YBW, CFR, and CGV, and MATα strains AKP and BBI—also displayed elevated noise levels relative to their mean fluorescence. To focus on the general trend across the population, these outlier offspring were excluded and data was fitted to a linear model. The resulting slopes were –0.22 for GFP and –0.52 for RFP, indicating a persistent but weaker inverse relationship between mean expression and noise compared to that observed among the parental strains. Importantly, the excluded offspring mirrored the distinctive noise profiles of their respective parents (Figure 4C-D). To systematically assess the heritability of gene expression noise across all the strains, we performed a "noise residual" analysis. This approach allowed us to quantify whether a strain is more or less noisy than expected for its expression level, enabling meaningful comparisons across strains and identification of heritable differences in noise regulation independent of expression level. We computed noise residuals for both parent strains and their offspring, analyzing GFP (from MATa) and RFP (from MATα) separately. For each mating type, we averaged the noise residuals across all offspring derived from a given parent and plotted these averages against the corresponding parent's noise residual. In MATa strains, we observed a significant positive correlation between parent and offspring GFP noise residuals ($r = 0.55$, $p = 1.99 \times 10^{-7}$; Figure 4E), indicating that parents with unusually high or low noise tended to produce offspring with similarly deviant noise levels. A similar analysis in MATα strains revealed a significant positive correlation for RFP noise residuals ($r = 0.41$, $p = 1.57 \times 10^{-2}$; Figure 4F). Reinforcing the notion that gene expression noise, often considered a stochastic trait (Elowitz et al., 2002)can exhibit heritable properties following sexual reproduction.

Additional correlation analyses, including comparisons between offspring noise residuals (for both GFP and RFP) and different parental summary metrics—mean, minimum, and maximum—are presented in Supplementary Figure S4G. While the average of the two parental noise residuals consistently showed the strongest correlation with offspring residuals (GFP: $r = 0.22$, $p < 0.001$; RFP: $r = 0.17$, $p < 0.001$), these correlations remain relatively modest, suggesting that an additive model alone cannot fully explain the inheritance of noise. This observation aligns with prior findings by(Ansel et al., 2008), demonstrating that gene expression noise in Saccharomyces cerevisiae follows a complex genetic

architecture, shaped by additive, dominance, and epistatic effects rather than simple linear inheritance patterns in sexually produced offspring.

Inheritance of Expression Noise Reveals Both Parental and Non-Parental Patterns

The observed correlation between parental and offspring noise residual levels suggests that the tendency of a strain to exhibit high or low noise levels has a genetic component and can be inherited across generations. Similar to our findings for mean expression levels, we hypothesized that the inheritance of gene expression noise may involve both parental and non-parental contributions. To test this, we applied an approach analogous to that used for fluorescence intensity. Specifically, we generated heatmaps of GFP and RFP noise residuals among the offspring, organized by half-sibling groups—defined as all offspring sharing the same parent from a specific mating type. In each heatmap, MATa parents are arranged along the vertical axis and MATα parents along the horizontal axis, allowing direct comparison of offspring that share one parent. As with the fluorescence intensity trait, we observed consistent patterns in noise residuals among offspring of specific parental strains. Some strains repeatedly gave rise to offspring with similar noise residual levels—either for the marker they contributed (parental inheritance) or for the marker from the opposite mating type (non-parental inheritance). These cases are highlighted with black dashed rectangles in Figure 4G-H.

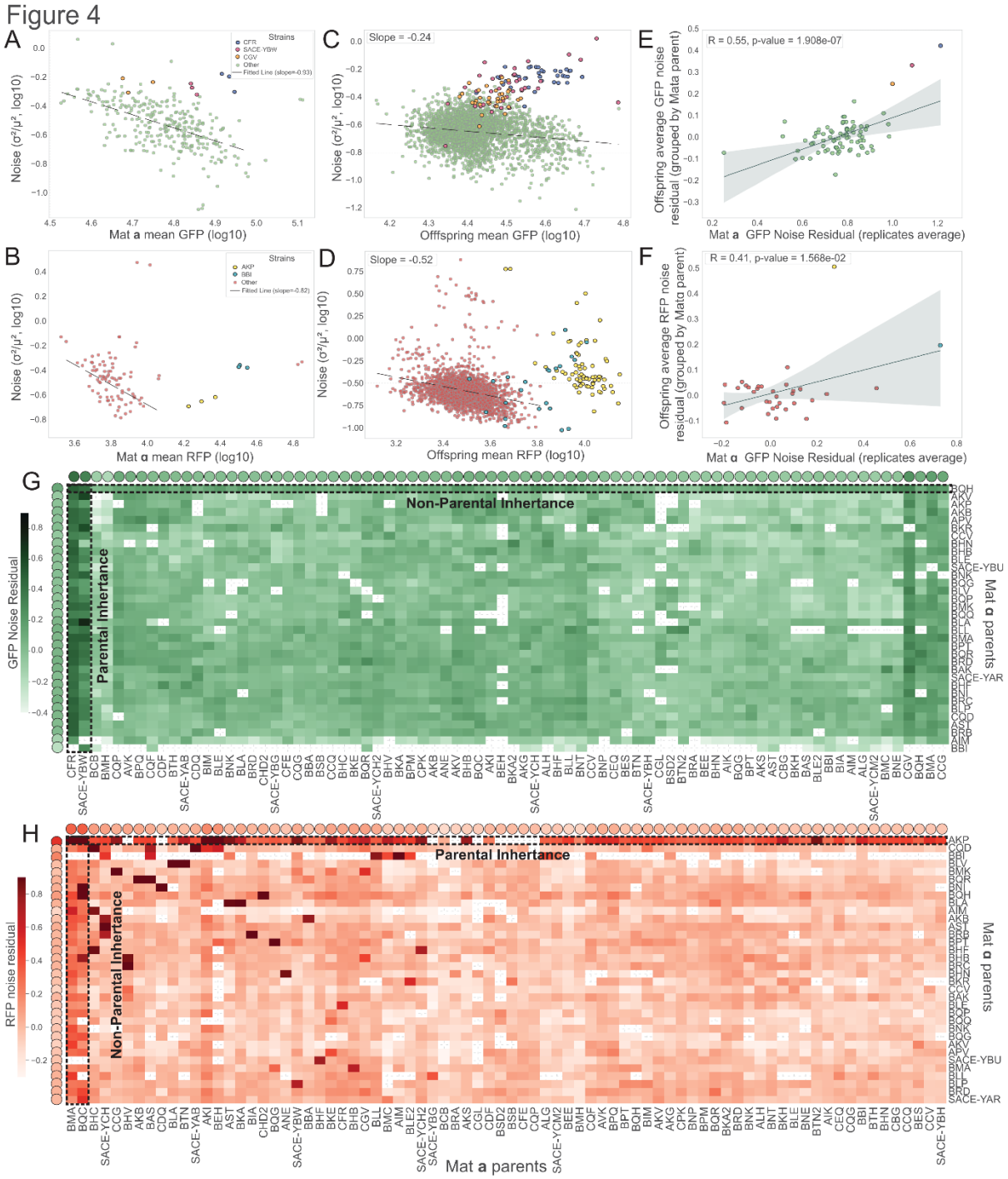

Figure 4 | Gene expression noise exhibits heritable properties following sexual mating and reveals parental and non-parental inheritance patterns. (A–B) Relationship between mean fluorescence level and expression noise in parental strains. (A) GFP-expressing MATa parents and (B) RFP-expressing MATα parents. Expression noise was quantified as the squared coefficient of variation (σ^2/μ^2) across FACS-measured populations. Noise and mean fluorescence are plotted on log–log axes. Linear fits excluding outliers (IQR method; see Materials and Methods) revealed negative slopes (–0.93 for GFP and –0.82 for RFP). Several parental strains deviated from this trend, including MATa strains SACE-YBW, CFR, and CGV, and MATα strains AKP and BBI, highlighted in color. (C–D) Relationship between mean fluorescence and expression noise in offspring. (C) GFP and (D) RFP expression levels are plotted against their respective noise values. Offspring derived from the high-noise parental strains shown in (A–B) are highlighted in matching colors. Linear fits excluding these high-noise offspring show persistent negative correlations (slopes: –0.24 for GFP; –0.52 for RFP). (E–F) Correlation between parental and offspring noise residuals. Residuals reflect deviations from the expected noise level given a strain's mean fluorescence (see Methods). A significant positive correlation was observed for both GFP (MATa parents, $r = 0.55$, $p = 1.91 \times 10^{-7}$; E) and RFP (MATα parents, $r = 0.41$, $p = 1.56 \times 10^{-2}$; F), indicating that noise is a heritable trait. (G-H) Heatmaps of GFP (G) and RFP (H) noise

residuals in diploid offspring, organized by half-sibling groups. MATa parents are arranged along the vertical axis and MATα parents along the horizontal axis. Consistent patterns of parental and non-parental inheritance are visible, with specific parental strains repeatedly producing offspring with similarly high or low noise residuals. Regions corresponding to parental and non-parental inheritance patterns are highlighted with black dashed rectangles.

Cross-Marker Correlation and Predictive Features of Gene Expression Noise Residuals in Offspring

When applying the model to predict the noise residuals of GFP and RFP expression in the offspring, we observed strong predictive performance for GFP, with an R^2 of 0.86 and a Pearson correlation of 0.93. Performance for RFP was more moderate, with an R^2 of 0.46 and a Pearson correlation of 0.68 (see Supplementary Figure 5C-D). Our analysis uncovered an interesting pattern: in both cases, the strongest predictor of one marker's noise was the noise residual of the other marker (Figure 5SC-D). This finding suggests that the propensity for expression noise is an intrinsic/extrinsic feature of the cell, rather than a marker-specific trait, and can influence multiple independent reporters. As in the models for mean expression, cell size features (forward scatter mean and standard deviation) emerged as important predictors for both GFP and RFP noise residuals. However, in contrast to the positive correlation observed for mean expression, higher forward scatter values were associated with lower predicted noise, suggesting that larger cells tend to exhibit more stable (less noisy) expression. In contrast to the results for mean expression levels, however, the noise residuals did not exhibit a strong parental effect. For GFP, 9 out of the 14 top predictive genes originated from the MATa parent (Figure 4A); for RFP, 8 out of 14 came from the MATα parent (Figure 4B). This more balanced distribution suggests that the regulation of noise is less tightly linked to parental origin and may instead reflect global, cell-intrinsic regulatory mechanisms. Among gene-level predictors, the strongest contributor to GFP noise residuals was YOL110W (SHR5), a gene encoding a palmitoyltransferase subunit. This enzyme complex modifies substrates such as Ras1p and Ras2p by adding a palmitoyl group via a thioester linkage, a modification essential for the membrane localization of Ras2p (Jung et al., 1995). For RFP noise residuals, the most important predictor was YKL047W (ANR2), a gene of unknown function that localizes to lipid droplets and is predicted to be palmitoylated. Its lipid-related localization and potential post-translational modification may indicate a broader connection between lipid metabolism and variability in gene expression (Figure 4).

Figure 5

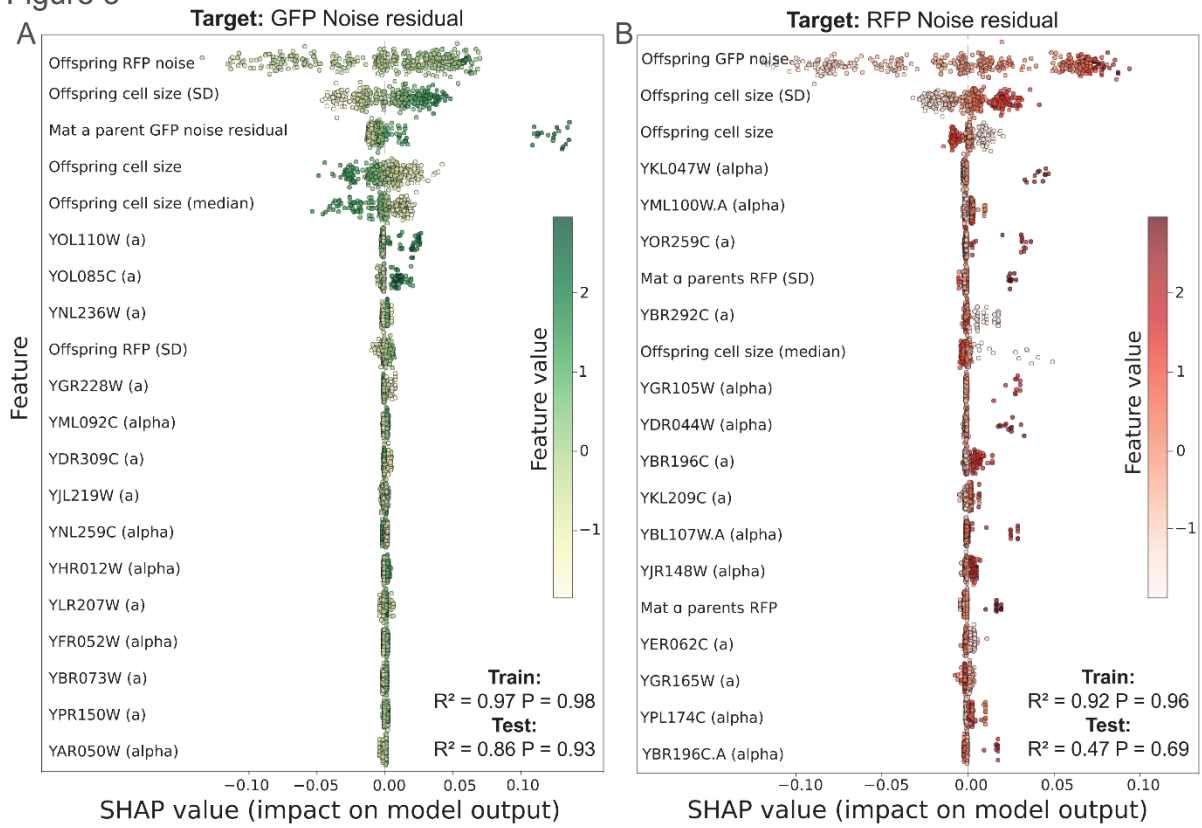

Figure 5 | SHAP summary plots reveal predictive features of noise residuals, including cross-marker effects and cell size dependence. (A) SHAP values for the model predicting GFP noise residuals in the offspring. Each point represents a sample, with its horizontal position indicating the SHAP value (i.e., the impact of that feature on the model's prediction). Points are colored by the normalized value of the feature. The model achieved $R^2 = 0.86$ and Pearson correlation = 0.93 on the test set. (B) SHAP values for the model predicting RFP noise residuals in the offspring, plotted using the same conventions as in (A). This model achieved a lower performance, with $R^2 = 0.47$ and Pearson correlation = 0.69 on the test set. In both models, the strongest predictor of one marker's noise residual was the noise residual of the other marker, indicating that the tendency for gene expression noise is a cell-intrinsic trait affecting multiple independent reporters. Additionally, cell size features (forward scatter mean and standard deviation) emerged as important predictors for both GFP and RFP noise residuals, and were negatively correlated, suggesting that larger cells tend to exhibit lower expression noise.

Discussion

The inheritance of quantitative traits has been a central theme in genetics and evolutionary biology. While gene expression levels have long served as a model quantitative trait, the inheritance of expression noise—the cell-to-cell variability in gene expression—has received comparatively little attention, particularly in the context of sexually reproducing organisms. In this study, we leveraged a large half diallel cross panel comprising over 2800 diploid yeast offspring to trace inheritance patterns of two quantitative traits: fluorescent protein expression levels and their associated expression noise.

A unique feature of our experimental system is the ability to directly compare the phenotypes of the parental haploid cells—the gamete equivalents in yeast—with those of their diploid offspring. In multicellular organisms, and especially in humans, such a comparison is inherently impossible: gametes are highly specialized, transcriptionally reduced cells whose phenotypes cannot be directly measured in a way that predicts the traits of the resulting zygote. By contrast, in *S. cerevisiae*, haploids are fully

functional vegetative cells whose gene expression profiles can be quantified with the same precision as diploids. This property allows us to treat haploid parents as direct phenotypic representatives of the gametes that give rise to the offspring, enabling an unprecedented, side-by-side assessment of trait inheritance from gametes to zygotes. Moreover, because haploids express each allele in isolation and diploids express the same allele in combination with its partner from the other parent, our setup allows us to trace allele-specific interactions directly. This means we can distinguish how an allele behaves on its own (reflecting cis-regulatory influences) versus how its expression changes in the presence of a second allele (revealing trans effects or allele–allele interactions). Such direct comparisons are rarely feasible in other systems, but here they provide a powerful framework for dissecting the mechanistic basis of non-additive inheritance.

This framework is particularly valuable for traits such as gene expression noise, where both genetic and trans-acting influences can shape variability, and where direct parental–offspring and allele-specific comparisons can reveal patterns that would otherwise remain hidden. Previous studies have shown that gene expression noise is a genetically encoded and evolvable trait. For example, (Duveau et al., 2018)demonstrated that promoter sequence variation can influence both mean expression and noise, with measurable fitness consequences. Similarly, (You et al., 2019) showed that the capacity to buffer gene expression noise is itself genetically determined and can be shaped by natural selection. However, such conclusions were drawn from isogenic strains and did not assess inheritance through sexual reproduction. Here, we provide direct evidence that gene expression noise is heritable following sexual mating. This is a notable finding, as it demonstrates that stochastic variation—often considered a purely intrinsic or transient feature of gene regulation—can be genetically transmitted across generations. Under certain environmental or selective conditions, this form of variability could contribute to phenotypic diversity and adaptive potential.

In the course of our analysis, we also uncovered a strong positive correlation between cell size (FSC-A) and mean fluorescence for both GFP and RFP. This pattern is consistent with the expectation that larger cells contain more total protein, leading to proportionally higher reporter signal. To account for this confounding factor, we normalized all fluorescence measurements by FSC-A, which effectively removed most of the size dependence in mean expression. Intriguingly, our noise analysis revealed a different relationship: in unnormalized data, noise was negatively correlated with cell size, in agreement with theoretical predictions (Jędrak & Ochab-Marcinek, 2020) that larger cells tend to buffer stochastic fluctuations more effectively. This dual role of cell size—enhancing mean expression while suppressing noise—suggests that size variation within populations can influence both the average output of gene expression and the stability of that output at the single-cell level. The persistence of strain-specific deviations from these trends after normalization further indicates that genetic factors can modulate noise independently of size, opening avenues to dissect the interplay between cell physiology and genetic regulation of stochasticity. Additionally, we revealed that expression inheritance follows a non-additive pattern, with two distinct modes. In parental inheritance, where a parent influences the expression of

the marker it contributed; in non-parental inheritance, the parent affects the expression of the marker contributed by its mate. These results are consistent with and extend findings from(Tsouris et al., 2024), who showed that a large fraction of gene expression traits in *S. cerevisiae* exhibit non-additive inheritance, primarily driven by trans-regulatory variation. Our results further suggest that these trans effects not only influence mean expression but also extend into expression noise. This non-parental mode is particularly intriguing, as it echoes features of both paramutation and transvection. Paramutation, first described in maize (Brink, 1956) and later observed in other organisms such as mice (Chandler, 2007), is an epigenetic process involving trans interactions in which one allele induces a heritable change in the expression of its homologous counterpart—without altering the DNA sequence. (Arteaga-Vazquez & Chandler, 2010). This modified state can be passed on to the next generation, violating traditional Mendelian expectations. Similarly, in "trans-vection", a phenomenon best characterized in *Drosophila* (Duncan, 2002; Lewis, 1954), gene expression is modulated through physical pairing and interaction between homologous alleles. These trans-allelic interactions can influence enhancer-promoter communication or chromatin state, resulting in allele-specific effects. A conceptually related phenomenon is seen in hybrid male sterility in *Drosophila*, where maternally inherited cytoplasmic factors—such as mitochondria—influence male fertility without passing the corresponding nuclear genes (Hoffmann et al., 1990). This represents a case in which a parent shapes an offspring's phenotype without transmitting the nuclear genetic determinant, and underscores the broader point that inheritance can operate through indirect or trans-acting mechanisms.

Although we did not test whether the non-parental effects we observe persist across multiple generations, the consistent, directional influence of one parent on the expression of the other parent's allele—particularly evident in our dual-reporter system—strongly supports the involvement of a trans-acting regulatory mechanism.

Materials and methods

Yeast strains and growth media

All yeast strains used in this study were derived from wild-type *Saccharomyces cerevisiae* natural isolates, selected from a previously curated collection (Peter et al., 2018). Approximately 100 strains were chosen and genetically engineered by integrating a genomic construct at the *HO* locus (details regarding strains selection and engineering can be found in Strauss et al., 2024, bioRxiv. Briefly, the construct included a unique 25-bp barcode specific to each strain, a Cre-lox based system to facilitate barcode fusion following mating, a constitutively expressed fluorescent protein, one antibiotic resistance gene driven by a constitutive promoter, and a second resistance gene driven by a mating type–specific promoter (to select for haploids during strains construction). Both MATa and Matα strains were engineered with the same general construct design, with specific variations in the reporter and selection markers. Specifically, MATa strains expressed green fluorescent protein (GFP) under the

control of the *DTH3* promoter, carried a hygromycin resistance gene (Hyg) under the constitutive TEF2 promoter, and a Zeocin resistant gene (BleoR) under a MATa-specific promoter (STE2). In contrast, MATα strains expressed red fluorescent protein (RFP) under the DTH3 promoter, carried a nourseothricin resistance gene (NAT1) under the TEF2 promoter, and a kanMX gene (conferring G418 resistance) under a MATα-specific promoter (STE3) (Strauss et al., 2024, bioRxiv).

Media and antibiotics:

YPD: 10 g/L yeast extract, 20 g/L peptone, and 20 g/L glucose.(Sherman, 2002).

SD Comp: 1.7 g/L nitrogen base (without amino acids and ammonium sulfate), 1 g/L monosodium glutamic acid (MSG), 1.5 g/L amino acid mix, and 20 g/L glucose.(Sherman, 2002)

SD Comp + DOXY: As above, with 10 µg/mL doxycycline.

Antibiotic concentration and initials:

Hygromycin (Hyg) - 300µg/ml, Zeocin (Zeo) - 150µg/ml, Geneticin (G418) - 200µg/ml and Nourseothricin (Nat) - 100µg/ml

FACS measurements of fluorescent protein levels in parental strains

All haploid parental strains (80 MATa and 47 MATα) were taken out of the -80°C and inoculated into YPD media with corresponding antibiotic (50 µg/mL zeocin for MATa strains and 200 µg/mL geneticin for MATα strains) in a 96-well plate using pinners. Strains were grown overnight (30°C, with shaking). After reaching stationary phase (~1×10^8 cells/mL) the strains were diluted 1:50 into SD Comp with corresponding antibiotic (50 µg/mL zeocin for MATa strains and 200 µg/mL geneticin for MATα strains) and grown for another 4.5 hours at 30°C with shaking to reach mid-log phase. OD was measured using a plate reader (infinite 200, Tecan). Following this, cultures were diluted 1:10 in phosphate-buffered saline (PBS) containing 0.05 M EDTA to reduce cell clumping. Plates were shaken for 2 minutes to ensure uniform suspension before loading into the FACS plate reader. Each plate was prepared in three technical replicates. FACS analysis was performed using an Attune NxT Flow Cytometer equipped with 405-, 488-, 561-, and 638-nm lasers. GFP (BL1) was detected using excitation at 488 nm and emission collection through a 530/30 bandpass filter, while RFP (mCherry) was detected using excitation at 561 nm and emission collected through a 620/15 bandpass filter. A threshold was applied to forward scatter (FSC) at 7.0×10^3 to reduce background noise and capture relevant cellular events. For each well, data were collected for 50,000 cells, including FSC (as a proxy for cell size), side scatter (SSC; for granularity), and fluorescence intensities for GFP and RFP. Gating parameters were optimized to include only single, viable cells. Gating was performed using Kaluza software (Beckman Coulter) and the Attune NxT cytometer interface (Thermo Fisher Scientific). For gating parameters see supplementary Figure S6. The same procedure was applied to all diploid offspring strains, with the exception that offspring were grown in SD Comp medium supplemented with both antibiotics (Nat 100 µg/mL and Hyg 200 µg/mL) to minimize contamination and eliminating remaining haploids. In offspring, both GFP and RFP expression were measured simultaneously for each strain in a single well.

In total, fluorescence measurements were collected for 3,002 offspring. For a randomly selected subset of 800 strains, two independent technical replicates were measured to assess reproducibility and fluorescence consistency.

Systematic mating of parental haploids to generate diploid offspring

All haploid parental strains (80 MATa and 40 MATα) were taken out of the -80°C and inoculated into YPD media with corresponding antibiotic (300 µg/mL Hyg for MATa strains and 100 µg/mL NAT for MATα strains) in a 96-well plate using pinners. Strains were grown overnight (30°C, with shaking). After reaching stationary phase (~1 × 10^8 cells/mL), cultures were diluted 1:1000 into SD Comp medium supplemented with the same concentrations of Hyg and Nat, and incubated for 12 hours to reach mid-log phase. OD was measured using a plate reader (infinite 200, Tecan). To ensure equal representation of each mating type in the mating reactions, optical density (OD) measurements were used to calculate volumes corresponding to 5 × 10^5 cells from each parent. Mating reactions were conducted in 96-well plates (a total of 35) containing 130 µL of SD Comp medium per well, supplemented with 10 µg/mL doxycycline (Doxy) to induce barcode fusion. For each mating reaction, 5 × 10^5 cells from one MATa strain and 5 × 10^5 cells from one MATα strain were added to the well, bringing the final volume to 140 µL. Plates were gently shaken for 5 minutes to ensure uniform mixing, then incubated at 25 °C without shaking for 20 hours to allow mating and Cre-lox-mediated barcode fusion. To eliminate the remaining haploid (parental) cells that failed to mate, cells were diluted 1:50 into SD Comp medium supplemented with 10 µg/mL doxycycline, as well as 300 µg/mL Hyg and 100 µg/mL NAT). Under these selective conditions, only offspring resistant to both antibiotics survived. This double-antibiotic selection step was performed twice, with a 24-hour interval between repetitions, to ensure complete removal of parental haploid cells and robust selection of diploid offspring. Following selection, all plates containing the 3,002 offspring strains were preserved as 30% glycerol stocks at –80°C for subsequent analyses.

Identity verification of the offspring

To confirm the genetic identity of the offspring and rule out contamination or errors during the mating process, we performed a barcode verification test on 56 randomly selected diploid strains. Fused barcodes were analyzed using Sanger sequencing. We amplified barcode regions using PCR with the following primers: (F: 5′-CCATACGAGCACATTACGGG-3′, R: 5′-CTTGACTGAGCGACTGAGG-3′ for MATa parent) and (F: 5′-CAGCGGGATAGTGCGATTG-3′, R: 5′-CAGCGGGATAGTGCGATTG-3′ for MATα parent). PCR was done in 25ul final volume with 2ul of template DNA (genomic DNA after cell lysis in 20mM NaOH and boiling for 20 minutes). PCR program: Tm of 60°C, elongation of 15 seconds, ~20 cycles.

Construct sequencing

To evaluate whether cis-regulatory variation contributes to differences in fluorescent reporter expression among offspring, we sequenced the regions surrounding the integrated reporter genes (including the promoter and terminator) in 20 MATa and 16 MATα parental strains selected for their divergent fluorescence levels. PCR was used to amplify the integrated construct regions using the following primers: (F: 5′-CAGCGGGATAGTGCGATTG-3′, R: 5′-CCAATGCGGAGCATATACGCC-3′ for GFP (MATa)) and (F: 5′-TACCACTCTTGACGACACGGC-3′, R: 5′-TTCTAATACCTGGAATGCTGT TTTGC-3′ for RFP MATα) PCR was done in 25ul final volume with 1ul of template DNA (genomic DNA). PCR program: Tm of 66°C, elongation of 1 minute, ~30 cycles.

Calculation of expression noise and noise residuals

To quantify gene expression variability, we calculated expression noise for both GFP and RFP as the squared coefficient of variation (σ^2/μ^2), where σ is the standard deviation and μ is the mean fluorescence intensity. These values were computed per strain based on flow cytometry measurements, typically from ~25,000 cells per strain after quality gating for parental strains and 35,000 cells for offspring strains (see above). To characterize how noise scales with expression level, we plotted noise values against mean fluorescence on log–log axes for both parental and offspring populations. Outlier parental strains were identified by visual inspection of noise versus mean expression plots. Strains were classified as outliers if they showed clear and consistent deviation from the global trend, such as elevated noise at moderate or high expression levels, or reversal of the expected negative relationship. No formal statistical outlier detection was applied, as the observed deviations were biologically interpretable and visually robust. The regression model was fitted to the central population using the interquartile range (IQR) method—an unbiased and systematic approach that includes only data points falling between the 25th and 75th percentiles of the mean expression distribution. For the offspring population, the regression line was fitted after manually excluding offspring derived from parental strains previously identified as noise outliers, in order to focus on the general trend within the main population. Noise residuals were then defined as the vertical distance (in log-space) between each strain's observed noise value and the predicted value from the fitted regression line. This residual measures how much more or less noisy a strain is compared to what would be expected based on its expression level, enabling standardized comparisons across strains.

Data Analysis

All data analyses were performed in Python (v3.10) using a modular workflow encompassing preprocessing, statistical modeling, and visualization. Raw FACS data were initially gated using the Attune NxT cytometer interface (Thermo Fisher Scientific), based on forward scatter (FSC), side scatter (SSC), and fluorescence intensities for GFP and RFP. Gating parameters are detailed in Supplementary Figure S6. The gated data were then exported directly as CSV files for downstream analysis. For each strain, the dataset included the mean and standard deviation of GFP and RFP fluorescence across the

gated population (comprising approximately 25,000 cells per strain for haploid parents and 35,000 cells per strain for diploid offspring). Subsequent data analysis was conducted using the pandas library for data manipulation and numpy for numerical operations. Statistical analyses—including correlation tests and linear regression modeling—were performed using scipy and statsmodels. Data visualizations, including boxplots, heatmaps, and regression plots, were created using matplotlib and seaborn. All analyses were reproducible with version-controlled scripts.

Machine Learning models and Analysis

To investigate the relationship between fluorescent protein expression and transcriptomic profiles, we leveraged data from a previous pan-transcriptome analysis of approximately 1,000 natural *Saccharomyces cerevisiae* isolates, focusing on a subset of ~120 strains, which included 4,977 core genes and 1,468 accessory genes (Caudal et al., 2024). For each offspring, we considered multiple features: the expression levels of all natural genes in each of its two parents (~12,000 features total), parental GFP and RFP expression levels, genetic distance between the parents, zygosity level (Peter et al., 2018), fitness of both parents and the offspring (Strauss et al., 2024, bioRxiv), and forward scatter measurements (as a proxy for cell size), collected via FACS alongside fluorescence intensity. We evaluated multiple machine learning models for prediction: LightGBM (LGB) (Guolin Ke, 2017), XGBoost (Chen & Guestrin, 2016), GradientBoostingRegressor (Friedman, 2001) and CatBoostRegressor (Liudmila Prokhorenkova, 2018). We selected a CatBoostRegressor model trained using a two-phase learning procedure, where an initial model was used to warm-start a second, extended training phase. This approach, which we refer to as DoubleLearningCatBoost, consistently achieved the highest R^2 across cross-validation folds and held-out test sets and was therefore used for all final analyses. Model performance was assessed in two steps. First, we performed a group-aware train/test split, ensuring that all offspring from a subset of unique mating pairs were held out for testing. This prevented data leakage from genetically related strains. Then, within the training set, we applied K-fold cross-validation to optimize model hyperparameters and select the best-performing estimator. After training, we applied SHAP (SHapley Additive exPlanations) to compute feature importance scores. SHAP assigns an importance value to each input by quantifying its contribution to the model's prediction, allowing us to interpret how gene expression, genetic distance, fitness, and cell size influenced predictive outcomes.

Supplementary

Figure S1

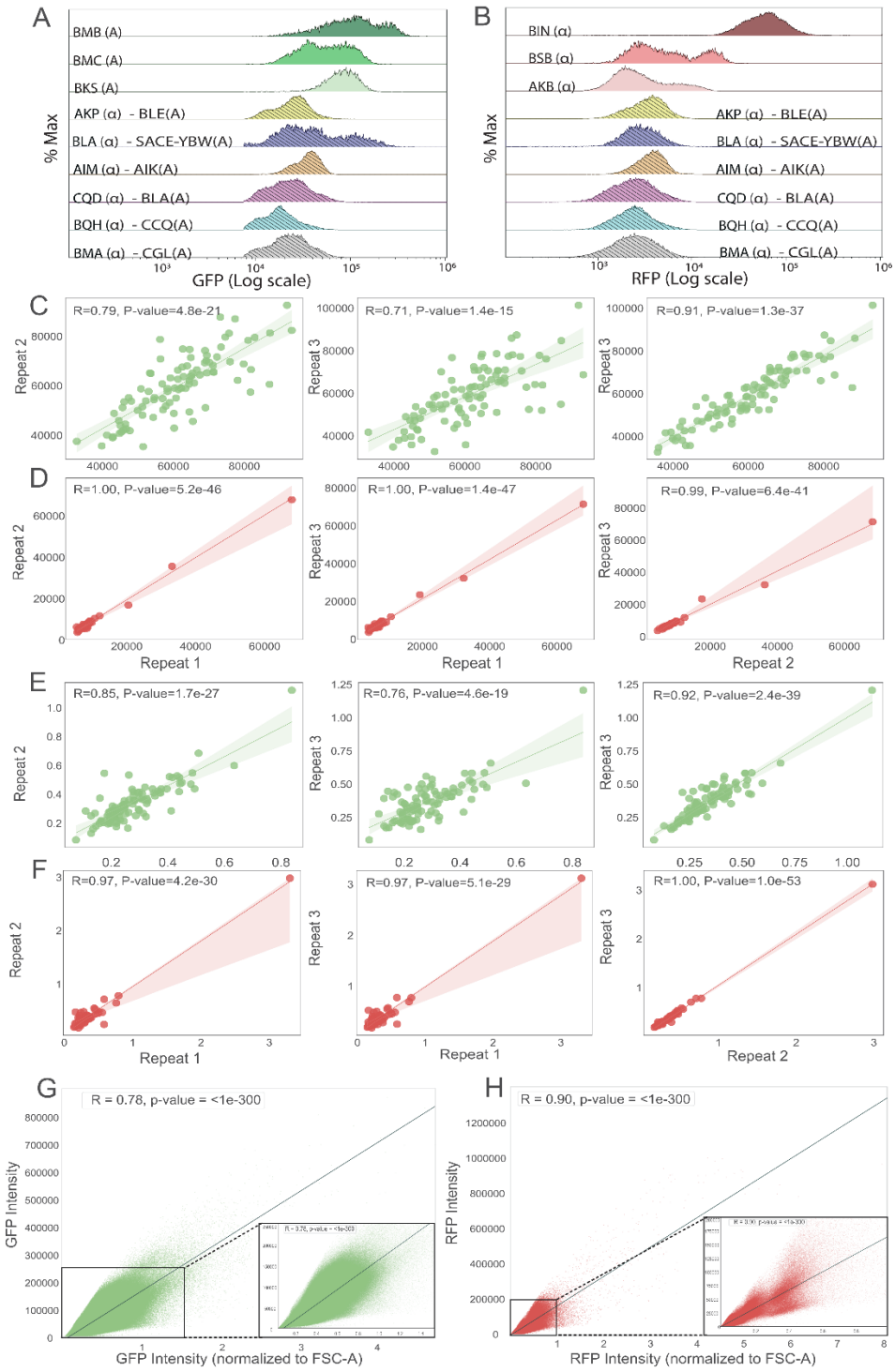

Figure S1 | Distributional variability in parents and offspring, strong correlation between biological replicates and effect of cell size normalization. (A–B) Distributions of fluorescence levels in individual cells in selected parental strains and randomly chosen offspring. For MATa parents, the strain with the highest GFP expression (BMB), a middle-expressing strain (BMC), and low expressing strain (BLS) were selected. For MATα parents, the highest RFP-expressing strain (BIN), a middle-expressing strain (BSB), and the lowest-expressing strain (AKB) were selected. In addition, six offspring were randomly selected for comparison. (E) GFP fluorescence distributions; (F) RFP fluorescence distributions. Fluorescence intensities are shown on a log scale (x-axis), and event counts are shown on the y-axis. Y-axis values represent "% Max"—the percentage of the maximum bin count across the distribution for each strain—allowing comparison between populations with different total

cell numbers. (C-D) Correlation of mean fluorescence intensity across three independent biological replicates for haploid parental strains. (A) GFP in MATa strains; (B) RFP in MATα strains. Each point represents the average expression of a single strain per replicate. High Pearson correlation coefficients indicate excellent reproducibility. (E–F) Correlation of fluorescence noise (defined as the squared coefficient of variation, CV^2) between replicates. (C) GFP noise in MATa strains; (D) RFP noise in MATα strains. Strong correlations confirm that noise estimates are consistent across experiments. (G–H) Correlation between raw fluorescence intensity and fluorescence normalized to cell size (FSC-A). (E) GFP in MATa strains; (F) RFP in MATα strains. Zoomed-in subplots highlight the main data cloud for clarity. Strong positive correlations ($r = 0.78$ for GFP and $r = 0.90$ for RFP; both $p < 10^{-300}$) indicate that normalization does not alter overall expression trends.

Figure S2

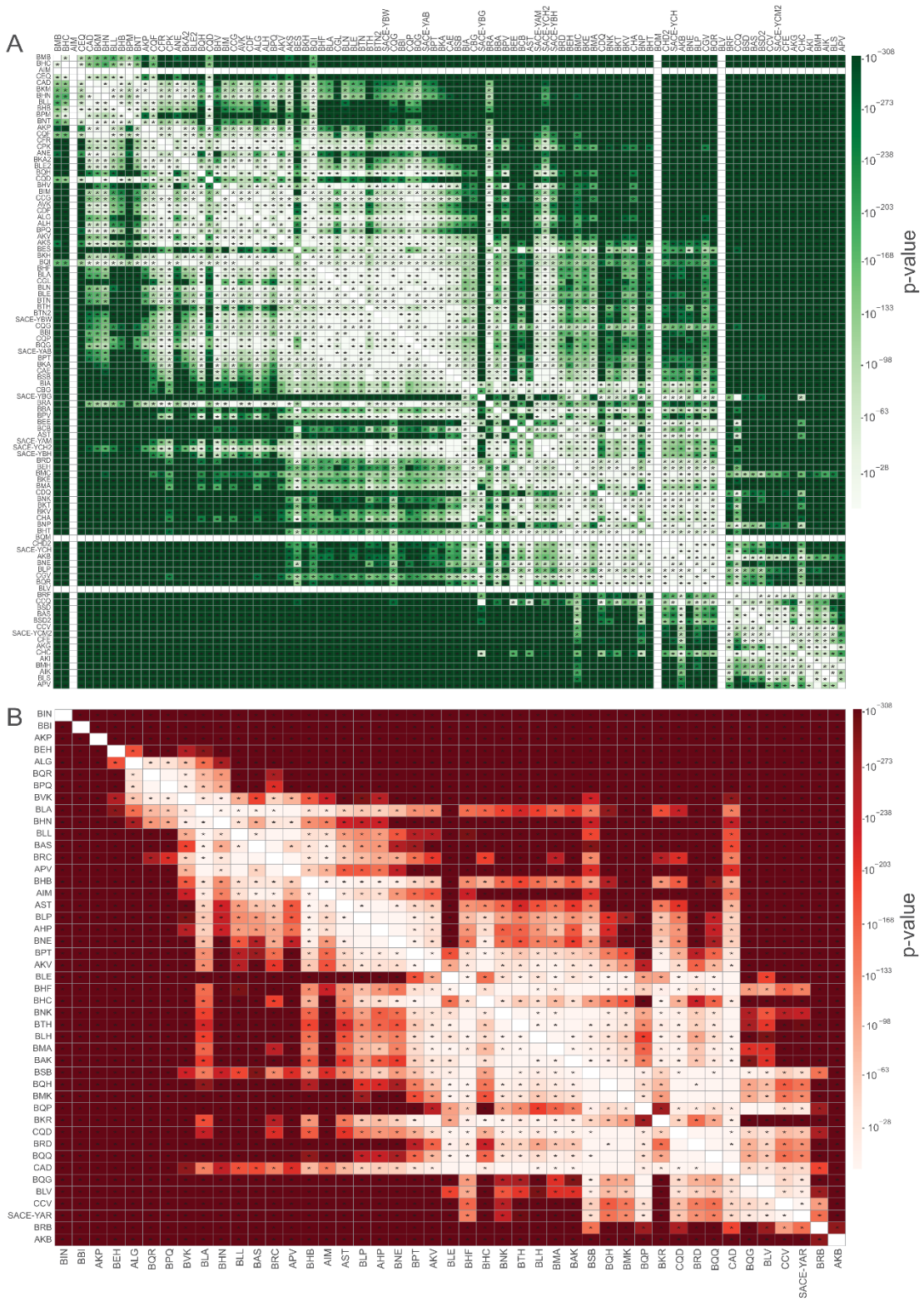

Figure S2 | Pairwise statistical comparison of fluorescence levels across all haploid parental strains. Heatmaps showing results from pairwise rank-sum tests comparing fluorescence distributions between all haploid parental strains within each mating type. Each cell in the matrix represents the p-value from a test comparing two strains, color-coded by significance level (scale bar at right). Asterisks within cells indicate significance ($p < 0.001$). (A) MATa strains and (B) MATα strains. Strains are ordered in descending order based on their mean fluorescence across three biological replicates. These results demonstrate that most strain pairs exhibit statistically significant differences in expression.

Figure S3

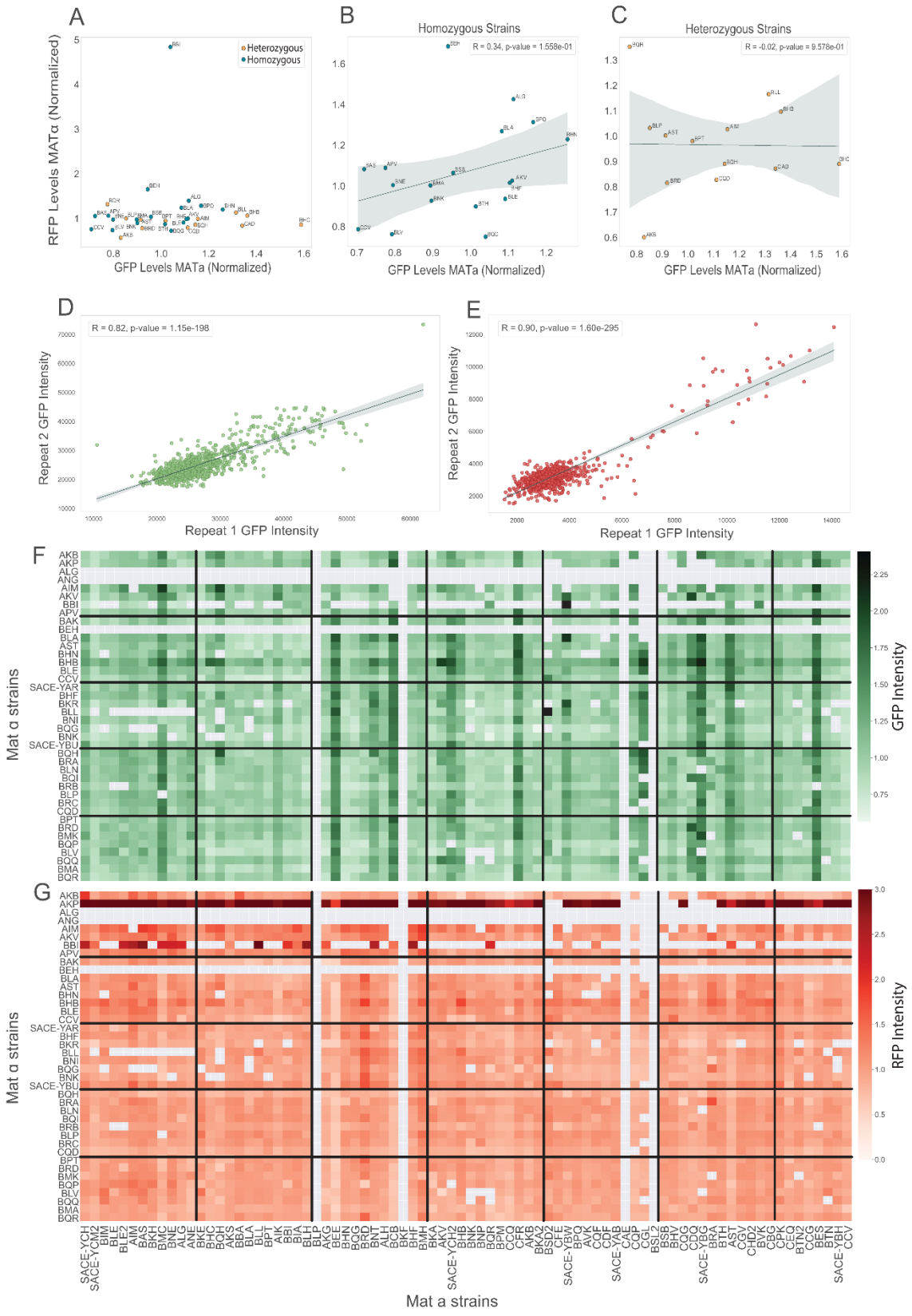

Figure S3 | Zygosity-based expression analysis, assessment of fluorescence measurement reproducibility and batch effects. (A–C) Zygosity-based comparison of GFP and RFP levels in 32 strains carrying both MATa and MATα mating types. (A) Scatter plot of normalized GFP and RFP expression, color-coded by zygosity class (homozygous = blue, heterozygous = orange). (B) Homozygous strains show a weak positive trend; (C) heterozygous strains exhibit no detectable correlation.

Neither trend was statistically significant. (D–E) Correlation of fluorescence measurements between two independent technical replicates for a subset of randomly selected 800 strains. (D) GFP expression correlation ($r = 0.82$, $p = 1.15 \times 10^{-198}$) and (E) RFP expression correlation ($r = 0.90$, $p = 1.6 \times 10^{-295}$) demonstrate strong reproducibility of the fluorescence measurements. (E–F) Heatmaps of GFP (E) and RFP (F) expression levels across 35 96-well plates, organized by plate layout. MATa parents are arranged along the x-axis and MATα parents along the y-axis. No evidence of batch effects or geographical biases was observed across plates.

Figure S4

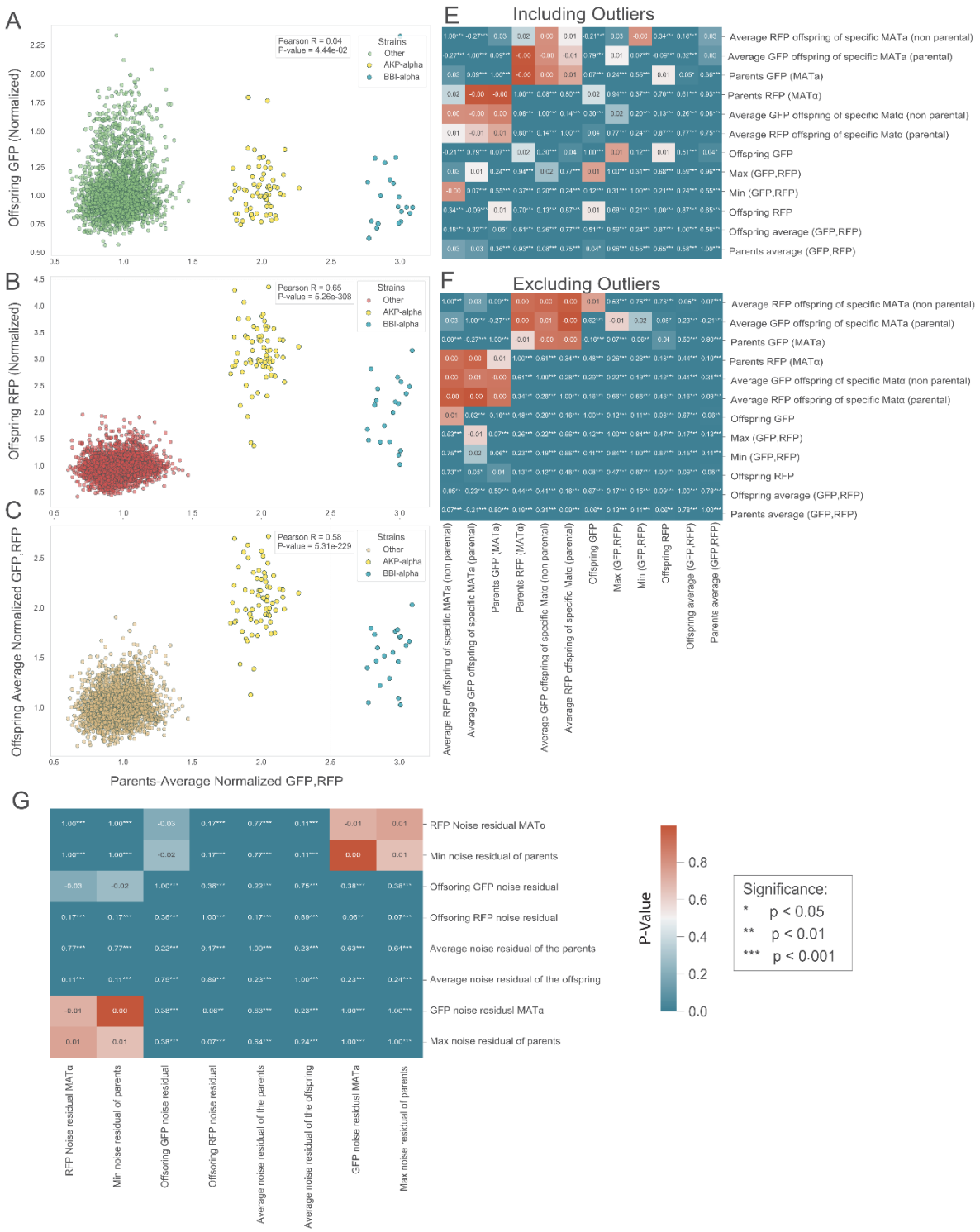

Figure S4 | Correlation analyses of parental and offspring fluorescence levels and noise residual, including and excluding outlier strains. (A–C) Relationship between offspring fluorescence and the parental average, with outlier strains

(offspring of MATα strains AKP and BBI) included. (A) Offspring GFP expression; (B) offspring RFP expression; (C) offspring average of GFP and RFP expression. Points corresponding to AKP- and BBI-derived offspring are highlighted in yellow and blue, respectively. (D–E) Correlation matrices showing relationships among various parental and offspring fluorescence inheritance metrics, including mean GFP, mean RFP, and average fluorescence, with (D) and without (E) inclusion of AKP- and BBI-derived offspring. Each cell displays the Pearson correlation coefficient (R), with statistical significance denoted by stars (*p < 0.05; **p < 0.01; ***p < 0.001), and cell color indicating the associated p-value. (F) Correlation matrix showing relationships between offspring noise residuals (GFP and RFP) and different parental inheritance metrics, including the parental noise residual, parental average, minimum, and maximum fluorescence values. As in D–E, Pearson correlation coefficient values are shown in each cell, stars indicate significance, and shading reflects the corresponding p-value.

Figure S5

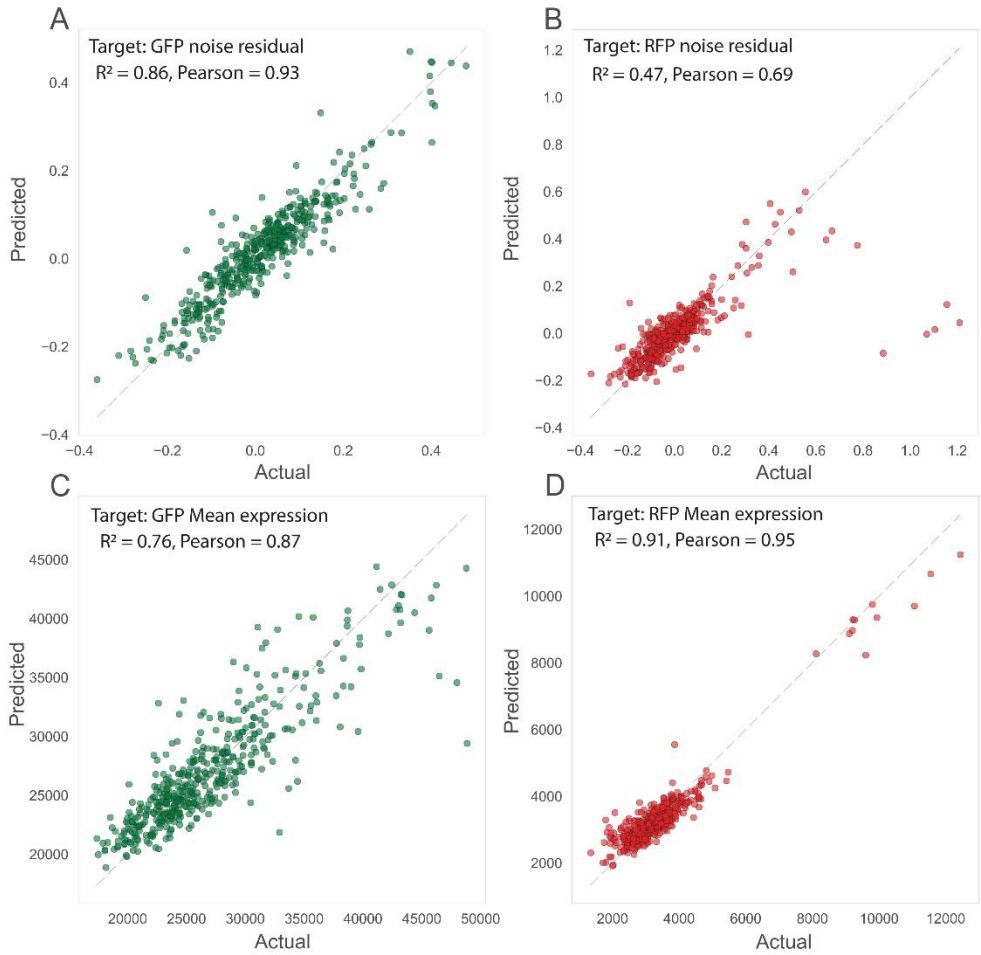

Figure S5 | High performance of machine learning models in predicting expression levels and noise residuals in diploid offspring. Predicted versus actual values are shown for four CatBoost regression models trained to predict offspring traits based on transcriptomic and genetic input features. (A) GFP mean expression levels in offspring. Predictions closely match actual values ($R^2 = 0.76$, Pearson = 0.87), indicating high model performance. (B) RFP mean expression levels in offspring. Predictions show very strong alignment with actual values ($R^2 = 0.91$, Pearson = 0.95), suggesting excellent predictive accuracy. (C) GFP noise residuals in offspring. The model achieves high predictive power ($R^2 = 0.86$, Pearson = 0.93), with predicted residuals closely tracking actual values. (D) RFP noise residuals in offspring. While predictive performance is lower than for other traits ($R^2 = 0.47$, Pearson = 0.69), the model still captures a meaningful signal in the data.

Figure S6

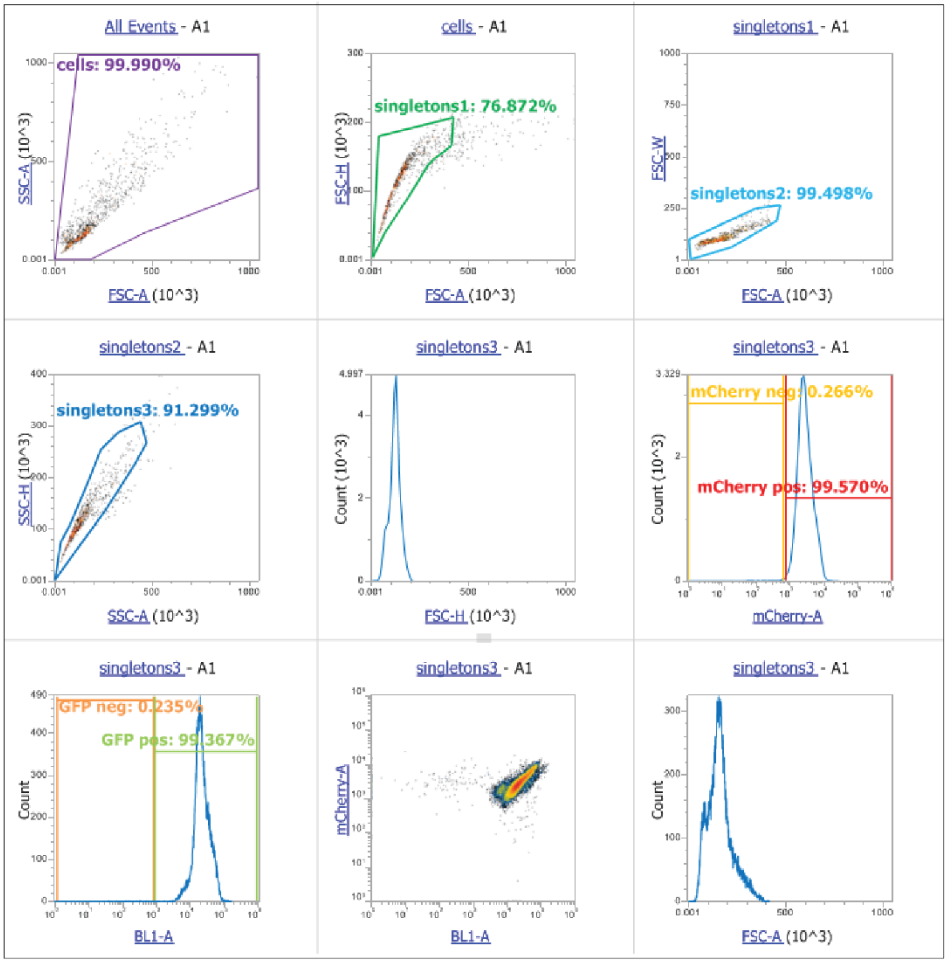

Figure S6 | FACS gating strategy for fluorescence measurement. Forward scatter (FSC) and side scatter (SSC) were used to identify the main cell population and exclude debris and aggregates. Sequential gating steps were applied to select single cells (singleton gates) and remove doublets based on FSC and SSC parameters. Fluorescence intensity of GFP and RFP was measured in the gated single-cell population using the BL1-A channel (GFP) and mCherry-A channel (RFP), recording the area (A) signal for each fluorophore.

Literature

Albert, F. W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. Nature Reviews Genetics, 16(4), 197–212. https://doi.org/10.1038/nrg3891

Ansel, J., Bottin, H., Rodriguez-Beltran, C., Damon, C., Nagarajan, M., Fehrmann, S., François, J., & Yvert, G. (2008). Cell-to-Cell Stochastic Variation in Gene Expression Is a Complex Genetic Trait. PLoS Genetics, 4(4), e1000049. https://doi.org/10.1371/journal.pgen.1000049

Arteaga-Vazquez, M. A., & Chandler, V. L. (2010). Paramutation in maize: RNA mediated trans-generational gene silencing. Current Opinion in Genetics & Development, 20(2), 156–163. https://doi.org/10.1016/j.gde.2010.01.008

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., & Barkai, N. (2006). Noise in protein expression scales with natural protein abundance. Nature Genetics, 38(6), 636–643. https://doi.org/10.1038/ng1807

Brem, R. B., Yvert, G., Clinton, R., & Kruglyak, L. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. Science, 296(5568), 752–755. https://doi.org/10.1126/science.1069516

Brink, R. A. (1956). A genetic change associated with the R locus in maize which is directed and potentially reversible. Genetics, 41(6), 872–889. https://doi.org/10.1093/genetics/41.6.872

Caudal, É., Loegler, V., Dutreux, F., Vakirlis, N., Teyssonnière, É., Caradec, C., Friedrich, A., Hou, J., & Schacherer, J. (2024). Pan-transcriptome reveals a large accessory genome contribution to gene expression variation in yeast. Nature Genetics, 56(6), 1278–1287. https://doi.org/10.1038/s41588-024-01769-9

Chandler, V. L. (2007). Paramutation: From maize to mice. Cell, 128(4), 641–645. https://doi.org/10.1016/j.cell.2007.02.007

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Cregg, J. M., Cereghino, J. L., Shi, J., & Higgins, D. R. (2000). Recombinant protein expression in Pichia pastoris. Molecular Biotechnology, 16(1), 23–52. https://doi.org/10.1385/MB:16:1:23

Duncan, I. W. (2002). Transvection effects in Drosophila. Annual Review of Genetics, 36, 521–556. https://doi.org/10.1146/annurev.genet.36.060402.100441

Duveau, F., Hodgins-Davis, A., Metzger, B. P., Yang, B., Tryban, S., Walker, E. A., Lybrook, T., & Wittkopp, P. J. (2018). Fitness effects of altering gene expression noise in Saccharomyces cerevisiae. eLife, 7, e37272. https://doi.org/10.7554/eLife.37272

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. Science, 297(5584), 1183–1186. https://doi.org/10.1126/science.1070919

Falconer, D. S., & Mackay, T. F. C. (1996). Introduction to Quantitative Genetics (4th ed.). Pearson Prentice Hall.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996). Life with 6000 genes. Science, 274(5287), 546–567. https://doi.org/10.1126/science.274.5287.546

Herskowitz, I. (1988). Life cycle of the budding yeast Saccharomyces cerevisiae. Microbiological Reviews, 52(4), 536–553. https://doi.org/10.1128/MMBR.52.4.536-553.1988

Hoffmann, A. A., Turelli, M., & Harshman, L. G. (1990). Factors affecting the distribution of cytoplasmic incompatibility in Drosophila simulans. Genetics, 126(4), 933–948. https://doi.org/10.1093/genetics/126.4.933

Lappalainen, T., Sammeth, M., Friedländer, M. R., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. Nature, 501(7468), 506–511. https://doi.org/10.1038/nature12531

Lewis, E. B. (1954). The theory and application of a new method of detecting chromosomal rearrangements in Drosophila melanogaster. The American Naturalist, 88(841), 225–239. https://doi.org/10.1086/281833

Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., et al. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature, 441(7095), 840–846. https://doi.org/10.1038/nature04785

Parapouli, M., Vasileiadi, A., Afendra, A.-S., & Hatziloukas, E. (2020). Saccharomyces cerevisiae and its industrial applications. AIMS Microbiology, 6(1), 1–32. https://doi.org/10.3934/microbiol.2020001

Peter, J., De Chiara, M., Friedrich, A., et al. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature, 556(7701), 339–344. https://doi.org/10.1038/s41586-018-0030-5

Rockman, M. V., & Kruglyak, L. (2006). Genetics of global gene expression. Nature Reviews Genetics, 7(11), 862–872. https://doi.org/10.1038/nrg1964

Ronald, J., & Akey, J. M. (2007). The evolution of gene expression QTL in Saccharomyces cerevisiae. PLoS ONE, 2(8), e678. https://doi.org/10.1371/journal.pone.0000678

Sherman, F. (1991). Getting started with yeast. In Methods in Enzymology (Vol. 194, pp. 3–21). https://doi.org/10.1016/0076-6879(91)94004-V

Sherman, F. (2002). Getting started with yeast. In Methods in Enzymology (Vol. 350, pp. 3–41). https://doi.org/10.1016/S0076-6879(02)50954-X

Smith, E. N., & Kruglyak, L. (2008). Gene–environment interaction in yeast gene expression. PLoS Biology, 6(4), e83. https://doi.org/10.1371/journal.pbio.0060083

Strauss, S. K., Golomb, R., Sheykhkarimli, D., et al. (2024). Quantitative genetics of natural S. cerevisiae strains upon sexual mating reveals heritable determinants of cellular fitness. bioRxiv. https://doi.org/10.1101/2024.02.12.579867

Tsouris, A., Brach, G., Schacherer, J., & Hou, J. (2024). Non-additive genetic components contribute significantly to population-wide gene expression variation. Cell Genomics, 4(1), 100459. https://doi.org/10.1016/j.xgen.2023.100459

You, S.-T., Jhou, Y.-T., Kao, C.-F., & Leu, J.-Y. (2019). Experimental evolution reveals a general role for the methyltransferase Hmt1 in noise buffering. PLOS Biology, 17(10), e3000433. https://doi.org/10.1371/journal.pbio.3000433