Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments

Christopher A Jackson[1,2,6], Dayanne M Castro[2,6], Giuseppe-Antonio Saldi[2], Richard Bonneau[1,2,3,4,5,7,*], David Gresham[1,2,7,*]

[1]Center For Genomics and Systems Biology, NYU, New York, NY 10008, USA
[2]Department of Biology, NYU, New York, NY 10008, USA
[3]Courant Institute of Mathematical Sciences, Computer Science Department, New York University, New York, NY 10003, USA
[4]Center For Data Science, NYU, New York, NY 10008, USA
[5]Flatiron Institute, Center for Computational Biology, Simons Foundation, New York, NY 10010, USA

[6, 7] These authors contributed equally to this work

*Correspondence: rb133@nyu.edu and dgresham@nyu.edu

Abstract

Understanding how gene expression programs are controlled requires identifying regulatory relationships between transcription factors and target genes. Gene regulatory networks are typically constructed from gene expression data acquired following genetic perturbation or environmental stimulus. Single-cell RNA sequencing (scRNAseq) captures the gene expression state of thousands of individual cells in a single experiment, offering advantages in combinatorial experimental design, large numbers of independent measurements, and accessing the interaction between the cell cycle and environmental responses that is hidden by population-level analysis of gene expression. To leverage these advantages, we developed a method for transcriptionally barcoding gene deletion mutants and performing scRNAseq in budding yeast (*Saccharomyces cerevisiae*). We pooled diverse genotypes in 11 different environmental conditions and determined their expression state by sequencing 38,285 individual cells. We developed, and benchmarked, a framework for learning gene regulatory networks from scRNAseq data that incorporates multitask learning and constructed a global gene regulatory network comprising 12,018 interactions. Our study establishes a general approach to gene regulatory network reconstruction from scRNAseq data that can be employed in any organism.

Introduction

Elucidating relationships between genes, and the products they encode, remains one of the central challenges in experimental and computational biology. A gene regulatory network (GRN) is a directed graph in which regulators of gene expression are connected to target gene nodes by interaction edges. Regulators of gene expression include transcription factors (TF) which can act as activators and repressors, RNA binding proteins, and regulatory RNAs. Identifying regulatory relationships between transcriptional regulators and their targets is essential for understanding biological phenomena ranging from cell growth and division to cell differentiation and development (Davidson, 2012). Reconstruction of GRNs is required to understand how gene expression dysregulation contributes to cancer and complex heritable diseases (Barabási et al., 2011; Hu et al., 2016).

Genome-scale methods provide an efficient means of identifying gene regulatory relationships. Efforts of the past two decades have resulted in the development of a variety of experimental and computational methods that leverage advances in technology and machine learning for constructing GRNs. Previously, we developed a method for inferring transcriptional regulatory networks based on regression with regularization that we have called the Inferelator (Bonneau et al., 2006; Ciofani et al., 2012). This method takes as inputs gene expression data and sources of prior information, and outputs regulatory relationships between transcription factors and their target genes that explain the observed gene expression levels. Subsequent work has enhanced this approach by selecting regulators for each gene more effectively (Madar et al., 2010), incorporating orthogonal data types that can be used to generate constraints on network

structure (Greenfield et al., 2013), and explicitly estimating latent biophysical parameters including transcription factor activity (Arrieta-Ortiz et al., 2015; Fu et al., 2011) and mRNA decay rates (Tchourine et al., 2018). We have successfully applied this approach to construct GRNs from gene expression data acquired from variation across time, conditions, and genotypes in microbes (Arrieta-Ortiz et al., 2015; Tchourine et al., 2018), plants (Wilkins et al., 2016), and mammalian cells (Ciofani et al., 2012; Miraldi et al., 2019).

Recently, single-cell RNA sequencing (scRNAseq) has exploded in popularity with the development of microfluidic systems for rapid encapsulation and labeling of thousands of cells in parallel. The DROP-seq system (Macosko et al., 2015) based on bead capture, and the inDrop (Zilionis et al., 2017) and 10x Genomics (Zheng et al., 2017) systems based on hydrogel beads, provide a facile means of generating RNA sequencing data for tens of thousands of individual cells. Although scRNAseq has primarily been used for defining cell types and states, this technology holds great potential for efficient construction of GRNs (Hwang et al., 2018). By combining genetic perturbation of transcriptional regulators using CRISPR/Cas9 with scRNAseq, mixtures of genetic perturbations can be assayed in a single reaction (Adamson et al., 2016; Dixit et al., 2016; Jaitin et al., 2016). This approach, known as Perturb-seq, presents a new opportunity for efficiently inferring GRNs from thousands of individual cells in which different regulators have been disrupted. There are considerable advantages in both scalability and detection of intra-sample heterogeneity with Perturb-seq, but quantifying the the effectiveness of CRISPR/Cas9 targeting in individual cells and distinguishing gene expression variability from noise inherent to mRNA undersampling in scRNAseq (Brennecke et al., 2013; Grün et al., 2014) present technical challenges. Computational methods to take advantage of scRNAseq data for inferring GRNs are under active development (Aibar et al., 2017; Chan et al., 2017; van Dijk et al., 2018). However, benchmarking these methods is difficult; in the absence of a known GRN, model performance is often estimated using simulated data (Chen and Mar, 2018), and issues regarding the appropriate experimental and computational approaches to GRN construction from scRNAseq data remain unresolved.

The budding yeast *Saccharomyces cerevisiae* is ideally suited to constructing GRNs from experimental data and benchmarking computational methods. Decades of work have provided a plethora of transcriptional regulatory data comprising functional and biochemical information (de Boer and Hughes, 2012; Teixeira et al., 2018). As a result, yeast is well suited to constructing GRNs using methods that leverage the rich available information and for assessing the performance of those methods by comparison to experimentally validated interactions (Ma et al., 2014; Tchourine et al., 2018). However, yeast presents several technical challenges for single cell analysis, and as a result scRNAseq methods for yeast reported to date (Gasch et al., 2017; Nadal-Ribelles et al., 2019) yield far fewer individual cells (~10^2) than are now routinely generated for mammalian studies (>10^4). The limitations of existing scRNAseq methods for yeast cells limits our ability to investigate eukaryotic cell biology as many signaling and regulatory pathways are highly conserved in yeast (Carmona-Gutierrez et al., 2010; Gray et al., 2004), including the Ras/protein kinase A (PKA), AMP Kinase (AMPK) and target of rapamycin (TOR) pathways (González and Hall, 2017; Loewith and Hall, 2011).

In budding yeast, the TOR complex 1 (TORC1 or mTORC1 in human) coordinates the transcriptional response to changes in nitrogen sources (Godard et al., 2007; Rødkær and Færgeman, 2014). Controlling this response are four major TF groups, which are regulated by diverse post-transcriptional processes. The Nitrogen Catabolite Repression (NCR) pathway, which is regulated principally by TORC1, consists of the TFs *GAT1*, *GLN3*, *DAL80*, and *GZF3* (Hofman-Bang, 1999), and is responsible for suppressing the utilization of non-preferred nitrogen sources when preferred nitrogen sources are available. Gat1 and Gln3 are localized to the cytoplasm until activation results in relocalization to the nucleus (Cox et al., 2000), where they then compete with Dal80 and Gzf3 for DNA binding motifs (Georis et al., 2009). The General Amino Acid Control (GAAC) pathway consists of the TF *GCN4* (Hinnebusch, 2005), and is responsible for activating the response to amino acid starvation, as detected by increases in uncharged tRNA levels. Gcn4 activity is translationally controlled by ribosomal pausing at upstream open reading frames in the 5' untranslated region (Mueller and Hinnebusch, 1986). The retrograde pathway, consisting of the TF heterodimer *RTG1* and *RTG3*, is responsible for altering expression of metabolic and biosynthetic genes in response to mitochondrial stress (Jia et al., 1997; Liao and Butow, 1993) or environmental stress (Ruiz-Roig et al., 2012). The Rtg1/Rtg3 complex is localized to the cytoplasm until activation, upon which they translocate into the nucleus (Komeili et al., 2000). The Ssy1-Ptr3-Ssy5-sensing (SPS) pathway (Ljungdahl, 2009), consists of the TFs *STP1* and *STP2*, and is responsible for altering transporter expression (Didion et al., 1998; Iraqui et al., 1999) in response to changes in extracellular environment. Stp1 and Stp2 are anchored to the plasma membrane until the SPS sensor triggers proteolytic cleavage of their anchoring domain and releases them for nuclear import (Andréasson and Ljungdahl, 2002).

Construction of GRNs based on the transcription factors in these pathways has had mixed success; the high redundancy of the NCR pathway has proven challenging to deconvolute (Milias-Argeitis et al., 2016). The GAAC pathway is more straightforward, although separating direct and indirect regulation remains difficult, even with high-quality experimental data (Mittal et al., 2017). As a result, a comprehensive GRN for nitrogen metabolism has remained elusive, despite successes in identifying genes that respond to changes in environmental nitrogen (Airoldi et al., 2016) and identification of post-transcriptional control mechanisms that underlie these changes (Miller et al., 2018).

Many signalling regulators involved in environmental response interact with cell cycle programs (Johnston et al., 1977; Talarek et al., 2017), including the TOR pathway (Zinzalla et al., 2007); however, how regulation of the mitotic cell cycle and environmentally responsive gene expression is coordinated is unknown. The regulation of nitrogen responsive gene expression in yeast is well-suited to the development of generalizable methods as the degree of TF redundancy, post-transcriptional regulation of TF activity, which precludes straightforward relationships between TF abundance and target expression, and multifactorial impact on gene expression, including intrinsic and extrinsic processes and stimuli, provide a tractable model system for addressing these challenges in higher eukaryotes.

Here, we have developed a method for scRNAseq in yeast using Chromium droplet-based single cell encapsulation (10x Genomics). We engineered TF gene deletions by precisely excising the entire TF open reading frame and introducing a unique transcriptional barcode that enables multiplexed analysis of genotypes using scRNAseq. We pooled 72 different strains, corresponding to 12 different genotypes, and determined their gene expression profiles in 11 conditions using scRNAseq analysis of 38,000 cells. We show that our method enables identification of cells from complex mixtures of genotypes in asynchronous cultures that correspond to specific mutants, and to specific stages of the cell cycle. Identification of mutants can be used to identify differentially expressed genes between genotypes providing an efficient means of multiplexed gene expression analysis. We used scRNAseq data in yeast to benchmark computational aspects of GRN reconstruction, and show that multi-task learning integrates information across environmental conditions without requiring complex normalization, resulting in improved GRN reconstruction. We find that imputation of missing data does not improve GRN reconstruction and can lead to prediction of spurious interactions. Using scRNAseq data, we constructed a global GRN for budding yeast comprising 12,018 regulatory interactions. We discover novel regulatory relationships, including previously unknown connections between regulators of cell cycle gene expression and nitrogen responsive gene expression. Our study provides a generalizable framework for GRN reconstruction from scRNAseq, a rich data set that will enable benchmarking of future computational methods, and establishes the use of large-scale scRNAseq analysis of multiplexed genotypes in yeast.

Results

Engineering a Library of Prototrophic, Transcriptionally-Barcoded Gene Deletion Strains

The yeast gene knockout collection (Giaever et al., 2002) facilitates pooled analysis of mutants using unique DNA barcode sequences that identify each gene deletion strain, but these barcodes are only present at the DNA level, precluding their use with scRNAseq. Therefore, we constructed an array of prototrophic, diploid yeast strains with homozygous deletions of TFs that control distinct regulatory modules: 1) NCR, 2) GAAC, 3) SPS-sensing, and 4) the retrograde pathway that coordinately control nitrogen-related gene expression in yeast. We engineered eleven different TF knockout genotypes, using six independently constructed biological replicates for each genotype. In addition, we constructed six biological replicates of the wild-type control in which we deleted the neutral HO locus. Genes were deleted using a modified kanMX cassette such that each of the 72 strains contains a unique transcriptional barcode in the 3' untranslated region (UTR) of the G418 resistance gene, that can be recovered by RNA sequencing (Figure 1, Supplemental Figure 1A). Homozygous diploids were constructed by mating to a strain containing the same TF knockout marked with a nourseothricin drug resistance cassette. On rich media plates, the 72 strains have an approximately wild-type growth; under nutritional stress, some TF knockouts exhibit growth advantages or disadvantages (Supplemental Figure 1B).

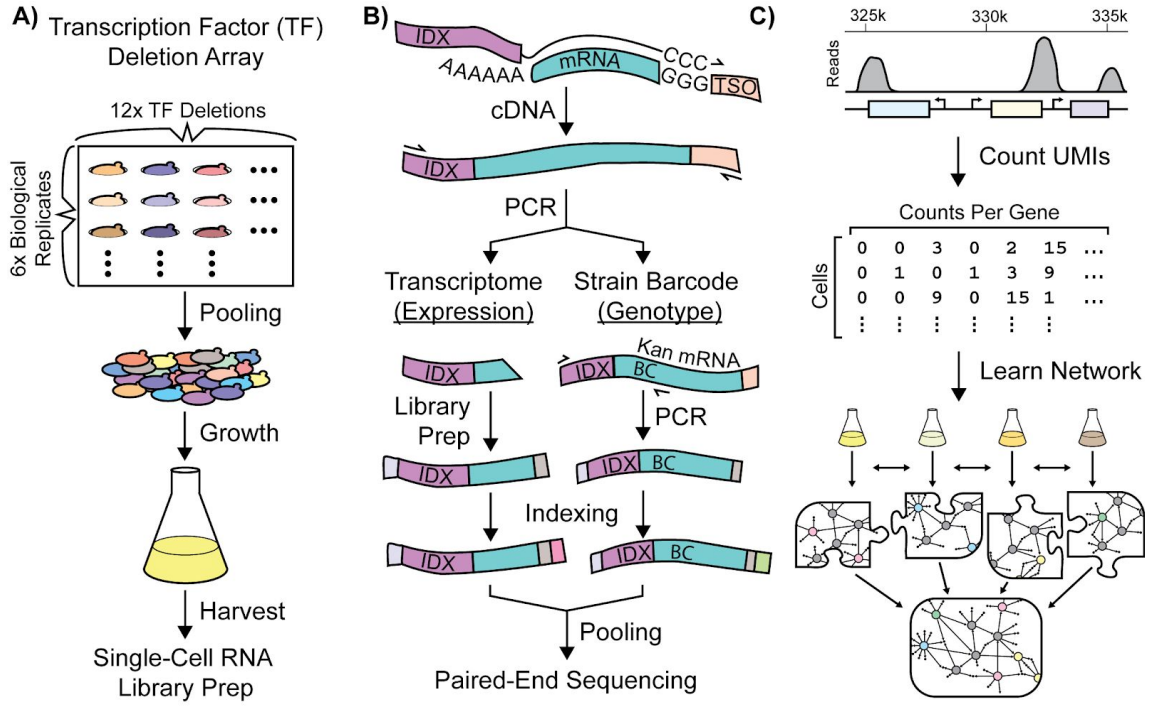


Figure 1: Single-Cell RNA-Seq Experimental Workflow in *Saccharomyces Cerevisiae*
Schematic workflows for: (A) Growth of a transcriptionally-barcoded pool of 11 nitrogen metabolism transcription factor (TF) knockout strains and a wild-type control strain each analyzed with six biological replicates (B) Synthesis in microfluidic droplets of single-cell cDNA with a cell-specific index sequence (IDX) attached to the oligo-dT primer, and a common template switch oligo (TSO). cDNA is processed for whole-transcriptome libraries, to quantify gene expression. In parallel, PCR products are amplified containing the genotype-specific transcriptional barcode (BC) encoded on the Kan^R antibiotic resistance marker mRNA, to identify cell genotype. Expression DNA libraries and PCR products are separately indexed for multiplexed sequencing (C) Processing of single-cell sequencing data into a Unique Molecular Identifier (UMI) count matrix which is used to learn a gene regulatory network using multi-task network inference from several different growth conditions

Table 1: Environmental Growth Conditions

Growth Condition	Abbrv.	Nitrogen Source	Carbon Source
Yeast Extract, Peptone, Glucose	YPD	YP	D-Glucose
YPD (Harvested after Post-Diauxic Shift)	DIAUXY	YP	D-Glucose
YPD + 200 ng/mL Rapamycin	RAPA	YP	D-Glucose
Yeast Extract, Peptone, Ethanol	YPEtOH	YP	Ethanol
Minimal Media (Glucose)	MMD	20 mM $(NH_4)_2SO_4$	D-Glucose
Minimal Media (Ethanol)	MMEtOH	20 mM $(NH_4)_2SO_4$	Ethanol
Nitrogen Limited Minimal Media (with Glutamine)	NLIM-GLN	0.8 mM L-Glutamine	D-Glucose
Nitrogen Limited Minimal Media (with Proline)	NLIM-PRO	0.8 mM L-Proline	D-Glucose
Nitrogen Limited Minimal Media (with NH_4)	NLIM-NH_4	0.8 mM $(NH_4)_2SO_4$	D-Glucose
Nitrogen Limited Minimal Media (with Urea)	NLIM-UREA	0.8 mM Urea	D-Glucose
Carbon Starvation	CSTARVE	1 mM $(NH_4)_2SO_4$	None

Environmental growth conditions are listed with their respective nitrogen and carbon sources. Yeast Extract + Peptone (YP) is a rich, complex nitrogen source. YP + Dextrose [YPD] is standard yeast rich media. Minimal media contains a standard base of trace metals, vitamins, and salts. All cultures were harvested 4 hours after inoculation, except for the culture harvested after the diauxic shift [DIAUXY], which was harvested 10 hours after inoculation. Rapamycin was added to YPD in the [RAPA] culture 30 minutes prior to harvest. Specific media formulations are listed in Supplemental Table 4.

Single-Cell RNA Sequencing of Pooled Libraries in Diverse Growth Conditions

ScRNAseq in yeast presents several challenges: cells are small (40-90 μm^3), enclosed in a polysaccharide-rich cell wall, and contain fewer mRNAs per cell (40k-60k) than higher eukaryotes. We developed and validated a protocol using the droplet-based 10x genomics chromium platform, and it used it to perform scRNAseq of the pool of TF knockouts in eleven growth conditions that provide a range of metabolic challenges (Table 1). In addition to variable nitrogen sources in minimal media with excess [MM] and limiting [NLIM-NH_4, NLIM-GLN, NLIM-PRO, NLIM-UREA] nitrogen, some conditions result in fermentative metabolism of glucose in rich media [YPD], and inhibition of the TOR signaling pathway in rich media by the small molecule rapamycin [RAPA]. We also studied conditions that require respiratory metabolism of ethanol in rich [YPEtOH] and minimal media [MMEtOH], and in rich media after sugars had been fully metabolized to ethanol and cells have undergone the diauxic shift [DIAUXY]. We tested two different starvation conditions, carbon [CSTARVE] and nitrogen starvation [NSTARVE]; however, the latter condition did not pass quality control during single-cell transcriptome library preparation and was discarded.

Cells from the eleven different conditions were sequenced and processed using cellranger (10x genomics) and our custom analysis pipeline (fastqToMat0), yielding a digital expression matrix (Supplemental Data 1) in which each cell is annotated with the environmental growth condition and genotype. Genotype-specific barcodes facilitate identification and removal of droplets that have multiple cells (doublets) by determining cell IDs that have more than one annotated genotype. Using our pool of 72 strains, we detect and remove 98.5% of doublets. PCR artifacts and duplicates are removed using Unique Molecular Identifiers (UMIs) (Kivioja et al., 2011) to quantify gene expression as unique transcript reads (UMI counts). Following sequence processing, quality control, removal of doublets, and assigning metadata, we recovered 83,703,440 transcript counts from a total of 38,225 individual cells.

To initially assess the quality of our data, we examined the expression of genes that are characteristic of different metabolic states. Consistent with our expectations, the core fermentative (anaerobic) genes *PDC1* and *ENO2* are expressed in cells in fermenting culture conditions only, and the core respirative (aerobic) gene *ADH2* is expressed in cells in respiring culture conditions (Figure 2A). The number of cells recovered varies by over an order of magnitude between conditions; stressful conditions of low nitrogen have lower cell yields overall. The yeast stress response is linked to increased resistance to zymolyase digestion (Nagarajan et al., 2014), which may be reflected in decreased cell yield during single-cell sequencing. Each of the 72 strains is found in each of the 11 conditions, although the number of each strain and genotype varies by environmental condition (Supplemental Figure 2A), and some strains are disproportionately affected. However, the number of transcripts per cell is generally equivalent between strains even when they differ in representation within libraries (Supplemental Figure 2B). By contrast, we find that total transcript counts per cell are highly linked to environmental growth condition (Supplemental Figure 2C), which is consistent with decreased total transcriptome pool size in suboptimal conditions (Athanasiadou et al., 2019).

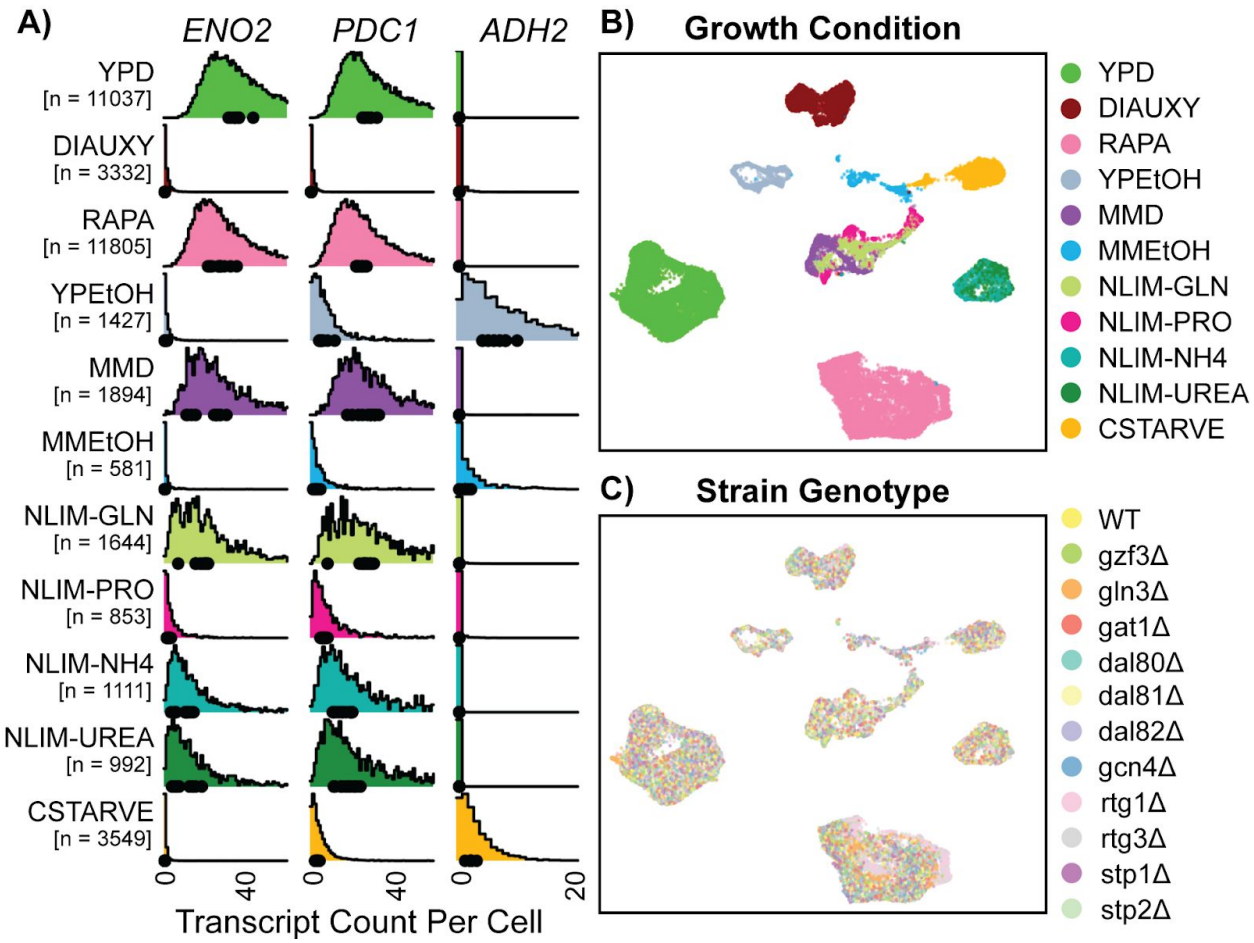


Figure 2: Gene expression of single Yeast Cells Cluster Based on Environmental Growth Condition (A) Normalized density histograms of raw UMI counts of the core glycolytic genes *ENO2* and *PDC1*, and the alcohol respiration gene *ADH2* in each environmental growth condition. Mean UMI count for each of the 12 different strain genotypes within each growth condition are plotted as dots on the X axis. (B-C) Uniform Manifold Approximation and Projection (UMAP) projection of log-transformed and batch-normalized scRNAseq data. Axes are dimensionless variables V1 and V2 with arbitrary units, here omitted. Individual cells are colored by environmental growth condition (B) or by strain genotype (C). Growth conditions are abbreviated as in Table 1.

For cells growing in rich medium (YPD) we recover a median of 2,250 unique transcripts per cell, from a median of 695 distinct genes, indicating a capture rate of approximately 3-5% of total transcripts from each cell. The strain genotype does not strongly influence transcript counts per cell (Supplemental Figure 2D). Mapping the digital expression matrix into two-dimensional space with a Uniform Manifold Approximation and Projection [UMAP] results in clear separation of individual cells into groups based on environmental condition (Figure 2B). Cells from different minimal media or nitrogen-limited growth conditions localize near each other, and cells grown in different rich nitrogen sources are clearly separate from each other. Within environmentally-determined grouping there appears to be no strong ordering by genotype (Figure 2C). Interactive figures are provided (http://shiny.bio.nyu.edu/YeastSingleCell2019/) facilitating exploration of expression levels for all genes.

The Mitotic Cell Cycle Underlies Heterogeneity in Single Cell Gene Expression

To identify sources of gene expression differences between cells within environments, we clustered single cells within each environmental condition separately by constructing a Shared Nearest Neighbor graph (Xu and Su, 2015) and clustering using the Louvain method (Blondel et al., 2008). Genes with known roles in mitotic cell cycle are highly represented among the most differentially expressed genes between clusters (Supplemental Figure 3A). Overlaying the expression of three of these genes (*PIR1*, *DSE2*, and *HTB1*/*HTB2*) on UMAP plots illustrates cell cycle effects on single cell gene expression (Figure 3A and Figure 3B). *PIR1* expression, a marker for early G1 (Spellman et al., 1998), is diagnostic of a distinct cluster. *DSE2* is expressed only in daughter cells (Colman-Lerner et al., 2001), which allows daughter cells in G1 to be distinguished from mother cells in G1. Cells that have high expression of the histone 2B genes, which are upregulated in S-phase (Eriksson et al., 2012), are localized together in the UMAP plots (Figure 3B).

For each cluster of cells within a growth condition we plotted the proportion of cells belonging to each TF deletion genotype, and the mean expression of several cell cycle genes (Figure 3B). Some clusters predominantly contain cells from a single TF deletion genotype; for example, cells deleted for *GLN3* (*gln3Δ*) form a separate cluster in YPD and RAPA conditions, as do cells deleted for one of the RTG heterodimer components (*rtg1Δ* and *rtg3Δ*). However, differences in expression due to genotype do not appear to be a primary source of expression differences within conditions, as most clusters show a uniform distribution of genotypes (Figure 3B, Supplemental Figure 3B). Similarly, we do not find that differences in expression of metabolic genes underlie overall differences in expression (e.g. *HXK2*) suggesting that the yeast metabolic cycle (Silverman et al., 2010; Tu et al., 2005) is not readily identifiable in single cells using scRNAseq.

Deletion of Transcription Factors Causes Gene Expression Changes that Differ Between Growth Conditions

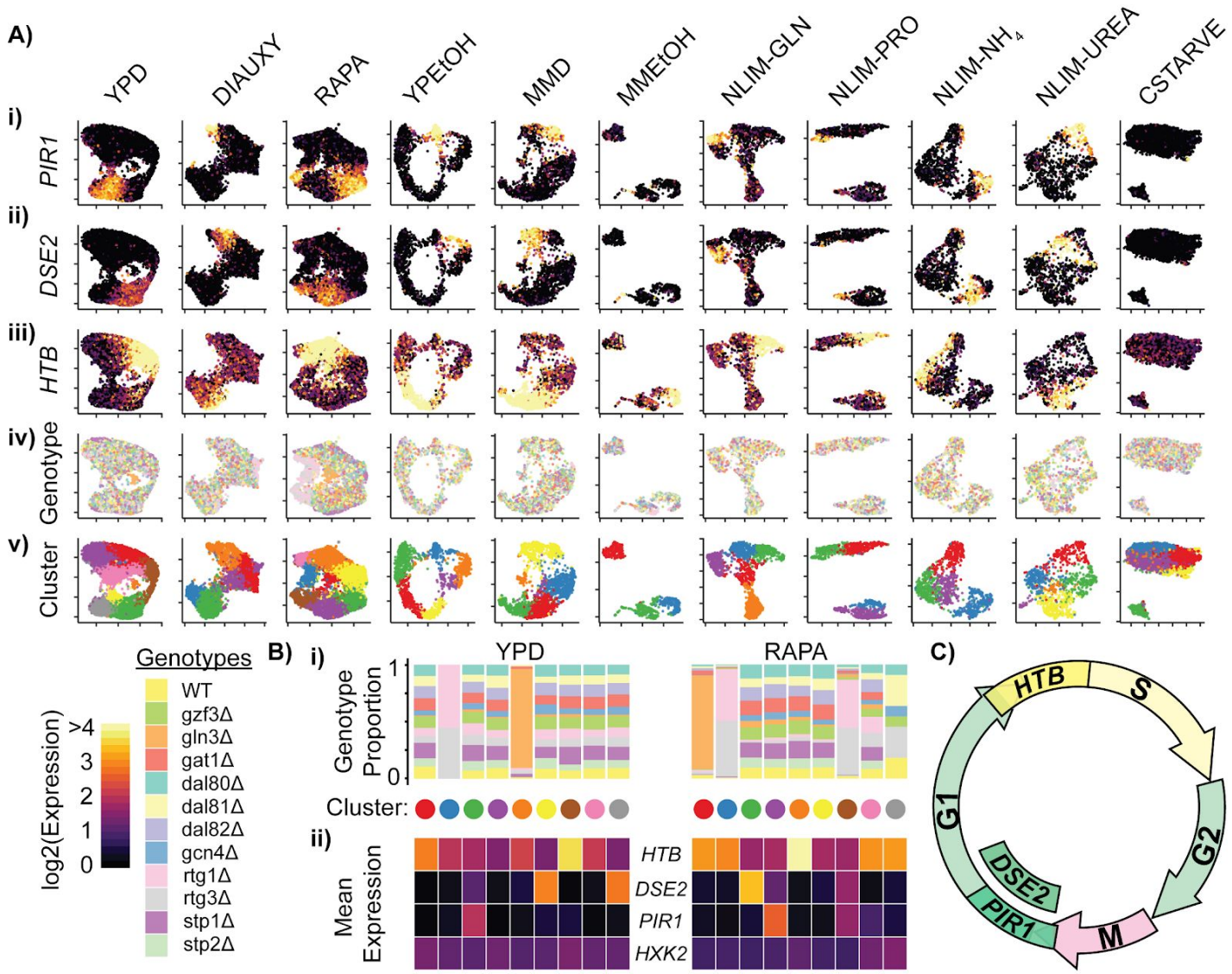


Figure 3: Cells Within Conditions Cluster According to Cell Cycle Genes (A) Cells from each growth condition were separately normalized and transformed into 2-dimensional space using UMAP. The log-transformed, normalized expression for each cell of (i) the G1-phase specific marker *PIR1*, (ii) the G1-phase daughter-cell specific marker *DSE2*, (iii) the S-phase specific marker histone 2B (*HTB*) is shown; (iv) the genotype and (v) the cluster membership of each cell. (B) Summary of clustered single cell expression within the YPD and RAPA growth conditions (i) Proportion of cells from a specific strain genotype within each cluster (ii) The mean log-transformed, normalized expression of the G1- and S-phase marker genes, as well as a hexokinase gene *HXK2* for each cluster (C) Schematic of the mitotic cell cycle with expression of *DSE2*, *PIR1*, and *HTB* genes annotated.

To assess our ability to determine differential gene expression between TF knockout strains, we examine the expression of genes known to respond to nitrogen signalling. *GAP1* (General Amino acid Permease) is a transporter responsible for importing amino acids under conditions of nitrogen limitation; *GAP1* expression is regulated by the NCR activators *GAT1* and *GLN3*, the NCR repressors *GZF3* and *DAL80* (Stanbrough and Magasanik, 1995), and potentially *GCN4* (Natarajan et al., 2001). We identify differing degrees of dysregulation of *GAP1* expression when these TFs are deleted (Figure 4A). The effect of deleting TFs varies by condition: *GAP1* is not expressed in YPD and its expression increases in nitrogen-limited media and in response to rapamycin. Deletion of *GAT1* results in decreased expression in nitrogen limiting media, but deletion of *GLN3* does not affect *GAP1* expression. By contrast, in the presence of rapamycin deletion of *GLN3* results in reduced *GAP1* expression. Deletion of *GCN4* only impacts *GAP1* expression in the presence of urea. *MEP2* and *GLN1* are also responsive to nitrogen TFs, and are dysregulated when certain TFs are deleted; expression of the glycolytic gene *HXK2* decreases when *GLN3*, *GCN4*, or *RTG1*/*RTG3* are deleted, but only in conditions of nitrogen limitation (Supplemental Figure 4A). These environmentally dependent impacts of genotype on gene expression demonstrate the importance of exploration of variable conditions for studying genotypic effects on expression.

A variety of statistical methods have been proposed and benchmarked for testing different expression of scRNAseq data (Soneson and Robinson, 2018); pairwise Welch's t-tests for differential expression are conceptually simple with few assumptions, and have high performance on scRNAseq data. Pairwise t-tests find few statistically significant (adjusted p-value < 0.05) differentially expressed genes (fold change > 1.5) that are due to deletion of a TF, but identify many differentially expressed genes between growth conditions (Supplemental Figure 4B). Our experimental design allows single-cell measurements to be collapsed into a total count (pseudobulk) measurement by summing counts across all cells that correspond to each of the six individual replicates of each genotype within a condition. When we analyze these data using standard approaches to RNAseq analysis (DESeq2) we detect several genes with significant (adjusted p-value < 0.05) differences in expression (fold change > 1.5) between wild-type and TF deletion strains (Figure 4B) that are consistent with known regulatory pathways. For example, in cells grown in rich media [YPD], we find 96 genes that are differentially expressed in TF deletion strains compared to wildtype (Figure 4C), many of which are annotated as functioning in amino acid metabolism and biosynthesis.

Optimal Modeling Parameters for Network Inference

Differential gene expression in a TF knockout strain is not sufficient evidence of a direct regulatory relationship as many significant changes in gene expression upon deleting a TF are indirect, and many direct effects may be subtle. Therefore, we constructed a gene regulatory network using the Inferelator, a regression-based network inference method which is based on three main modeling assumptions. First, we assume that Transcription Factor Activity (TFA) is a latent biophysical parameter that represents the effect of a TF binding to DNA and modulating its transcription activity (Arrieta-Ortiz et al., 2015; Fu et al., 2011). The TFA values are not

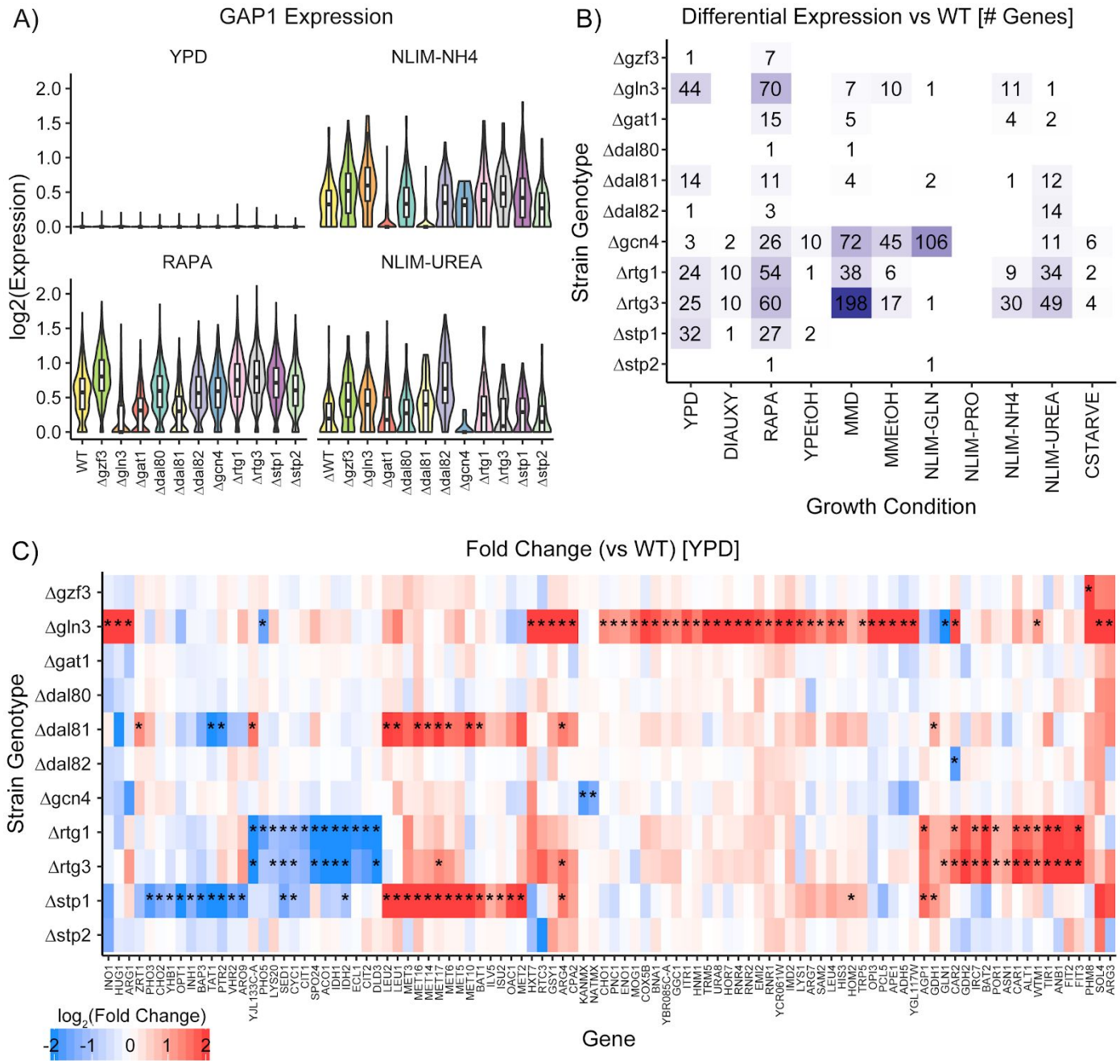


Figure 4: Impact of Deleting Transcription Factors on Gene Expression (A) Violin plots of the $\log_2$ batch-normalized expression of the general amino acid permease gene *GAP1* in YPD, RAPA, ammonium-limited media, and urea-limited media. (B) Count of differentially expressed genes in each combination of growth condition and strain genotype. Data were transformed to pseudobulk values by summing all counts for each the six biological replicates for each genotype and then analyzed for differential gene expression using DESeq2 [1.5-fold change; p.adj < 0.05]. (C) Log_2(fold change) of genes differentially expressed in TF knockout strains compared to wildtype, when grown in YPD. Asterisks denote statistically significant differences in gene expression [1.5-fold change; p.adj < 0.05].

directly measured, and instead must be estimated as a relative value based on prior knowledge of a regulatory network of TF and target relationships. This TFA estimation is essential as many TFs are post-transcriptionally regulated, or are expressed at levels that are not reliably detected by scRNAseq (Filtz et al., 2014). Second, we assume that expression of a gene can be described as a weighted sum of the activities of TFs (Bonneau et al., 2006) using an additive model in which activators and repressors increase or decrease the expression of targets linearly. Finally, we assume that each gene is regulated by a small number of TFs, and that regularization of gene expression models is required to enforce this biologically relevant property of target regulation. *Saccharomyces cerevisiae*, as a preeminent model organism in systems biology, has a well defined set of known interactions that are of considerably higher quality than is available for more complex eukaryotes providing a validated gold standard for testing model performance (Tchourine et al., 2018).

To evaluate the performance of data processing methods and model parameter selections within the Inferelator on scRNAseq data, we perform ten cross-validations using the existing gold standard network. During cross-validation, we infer a GRN using half of the gold standard target genes as priors, then evaluate performance based on recovery of TF-target gene interactions for gold standard interactions that are left out of the priors. We tested preprocessing and prior selection options by inferring networks using gene expression models that are regularized by best subset regression to minimize Bayesian Information Criterion (Arrieta-Ortiz et al., 2015; Greenfield et al., 2013) and quantified performance in predicting TF-target interactions using the area under the precision-recall curve (AUPR). As negative controls, we employed the same procedure after shuffling priors and after simulating scRNAseq data in which all variance is due to sampling noise. The negative control with shuffled priors establishes a random classifier baseline AUPR of 0.02; the negative control with simulated data establishes a circular recovery baseline AUPR of 0.06 (Figure 5A). Performance of the Inferelator on our scRNAseq data far exceeds these baselines, with a mean AUPR of 0.20. This performance from our single dataset is comparable to that of a GRN constructed from 2,577 experimental observations using bulk gene expression data (Tchourine et al., 2018).

The sparsity of data for each cell acquired using scRNAseq may negatively impact its utility in GRN construction. A commonly used technique to address missing data is data imputation. We tested the impact of several imputation packages on network inference: MAGIC (van Dijk et al., 2018), ScImpute (Li and Li, 2018), and VIPER (Chen and Zhou, 2018). Whereas these methods can enhance separation of gene expression states in low-dimensionality projections (Supplemental Figure 5A), we find that they are either ineffective or detrimental to network inference (Figure 5A). When the GRN is reconstructed from interactions selected at a precision threshold of 0.5, which takes into account how many interactions are correct according to the gold standard, no imputation method increases the number of recovered interactions compared to unmodified data. Data imputation with MAGIC increases the total number of confidently predicted (confidence > 0.95) interactions, but recovers fewer interactions that are correct according to the gold standard.

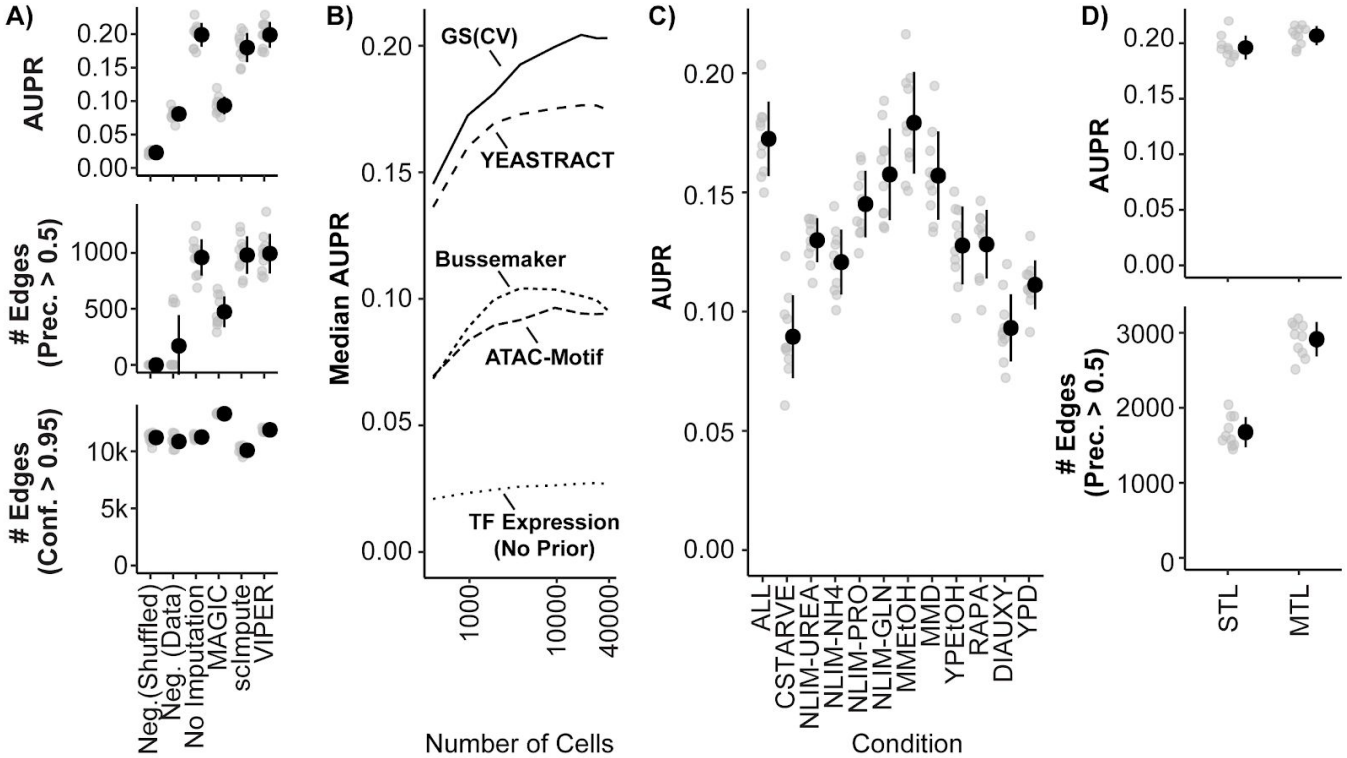


Figure 5: Model Performance and Impact of Data Imputation, Prior Selection, and Multitask learning on Network Inference using the Inferelator

(A) Model performance of Inferelator (TFA-BBSR) network inference after shuffling priors [Neg. Shuffled], on a simulated negative data set [Neg. Data], on the unaltered count matrix [No Imputation], and after imputing missing data from the count matrix using the MAGIC, ScImpute, and VIPER packages. Model performance is shown using area under the precision-recall curve [AUPR], as well as the number of network edges using a precision (> 0.5) cutoff, and the number of network edges using a confidence (> 0.95) cutoff. Each point plotted in gray is a separate cross-validation analysis, with mean +/- one standard deviation plotted in black (n=10). (B) Median AUPR after cross-validation (n=10) and resampling to different numbers of cells, for priors extracted from the gold standard [GS], the YEASTRACT database, Bussemaker et al, priors predicted from ATAC-seq data and motif searching, and no prior data. (C) AUPR of separate cross-validation network inference using cells from all growth conditions, or from individual conditions separately. Each cross-validation (n=10) was downsampled to the same number of cells. (D) AUPR and number of network edges per cross-validation (n=10) for single-task learning (STL), where all cells are included and inference is performed with single-task learning (TFA-BBSR), or multi-task learning (TFA-AMuSR), where all cells are included, but are separated into individual tasks. The YEASTRACT prior data is used for both methods, resulting in mean AUPR of 0.197 for STL and 0.209 for MTL.

Selection of Priors for Inference

Algorithms for network inference perform poorly when making predictions based only on expression data (Greenfield et al., 2010). Including prior knowledge of regulatory relationships and network topology improves model selection, and allows approximation of latent variables like TFA. Priors can be generated from regulatory interactions defined using methods such as chromatin immunoprecipitation sequencing (ChIP-seq) or analysis of transposase-accessible chromatin (ATAC-seq) and TF binding motifs, or from curated databases of interactions derived from literature. The source and processing of prior knowledge has a substantial effect on the size and accuracy of the learned network (Azizi et al., 2018; Siahpirani and Roy, 2017). We tested the impact on GRN reconstruction of prior data derived from literature, and from high-throughput experimental assays that encompass interactions between the entire yeast genome and the majority of known TFs (Figure 5B). The best performance is obtained using a curated set of known TF-gene interactions obtained from YEASTRACT (Teixeira et al., 2018). Generating priors using motif searching within open chromatin regions determined by ATAC-seq (Castro et al., 2019; Miraldi et al., 2019), and by modeling TF-DNA affinities in promoters (Ward and Bussemaker, 2008) provides a considerable improvement over GRN reconstruction from TF expression without priors, but have lower performance than priors derived from curated data.

Multi-task Learning Improves Network Inference and Enables Reconstruction of a Unified Gene Regulatory Network from Multiple Conditions

Numerous methods exist for integrating information across different conditions and experiments that aim to reduce technical variation while retaining biologically meaningful differences (Hicks et al., 2018; Leek et al., 2010). The appropriate approach to integrating scRNAseq data for the purpose of GRN reconstruction remains unknown. We find that when we separate data based on environmental condition and infer GRNs we obtain unique networks of differing quality (Figure 5C). Learning a single network from all conditions by first combining the data can be compromised by technical variability and imbalance in the number of cells between conditions. Furthermore, normalizing batches to equal transcript depth risks suppressing differences which are true biological variability. An alternative approach is to treat the cells from each environmental condition as separate tasks, and then jointly learn a network using a multitask learning (MTL) framework (Castro et al., 2019; Lam et al., 2016). We find that our multi-task network inference (MTL) procedure (Castro et al., 2019), which we named Adaptive Multiple Sparse Regression (AMuSR), improves the quality of the network inference and increases the size of the network recovered (Figure 5D). This is likely due to the retention of environmentally specific interactions that would otherwise be obscured using methods for normalizing data prior to GRN construction.

We constructed a global gene regulatory network using the YEASTRACT priors (as determined above) and our multi-task network inference (MTL) procedure. Eleven GRNs were jointly learned from each of the 11 environmental growth conditions; for each task a confidence score

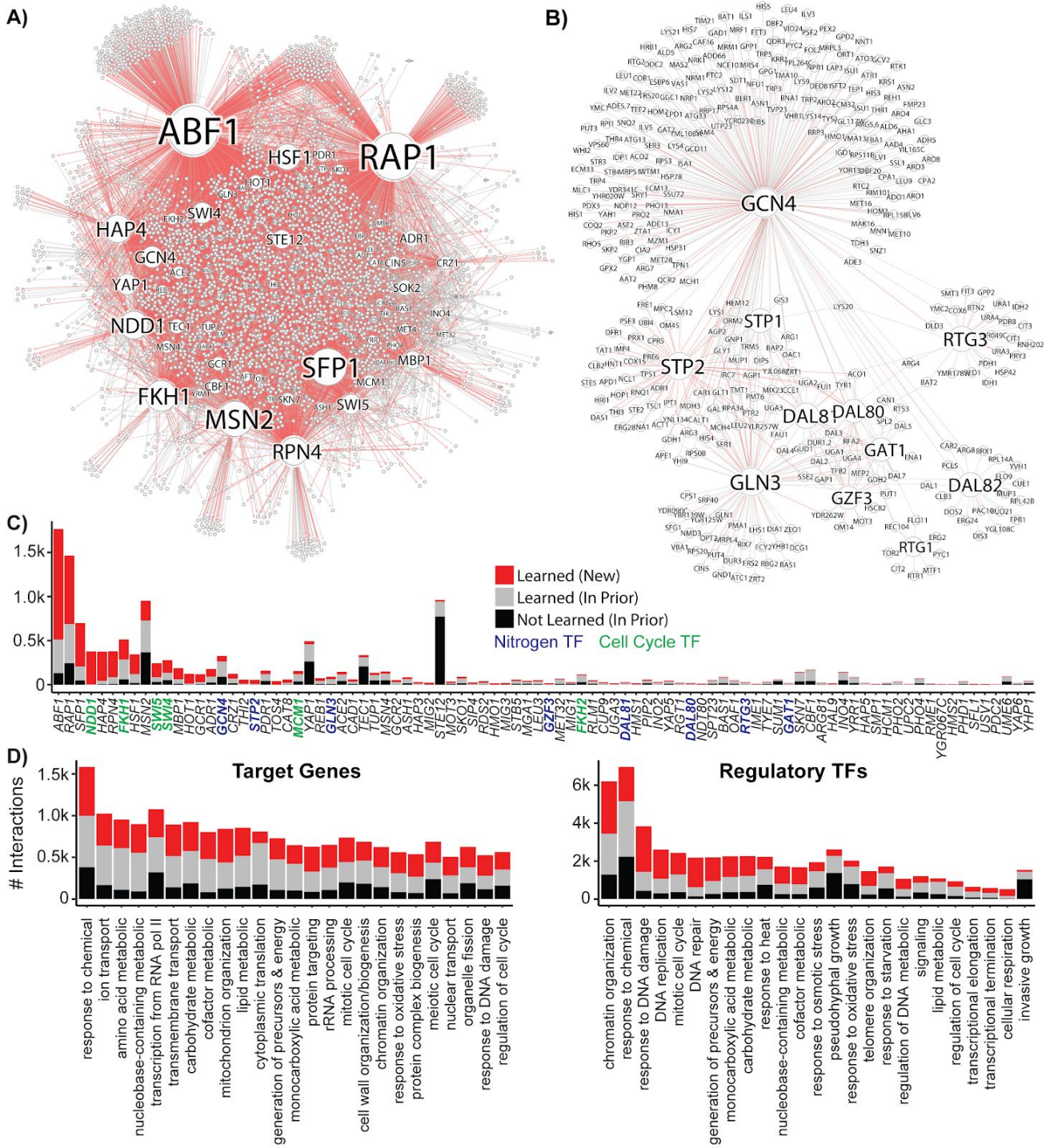


Figure 6: Reconstruction of a Gene Regulatory Network Identifies New Regulatory Relationships A network inferred from the single-cell expression data using multi-task learning and the YEASTRACT TF-gene interaction prior, with a cutoff at precision > 0.5. (A) Network graph with known interaction edges from the prior in gray and new inferred interaction edges in red (B) Network graph of the 11 nitrogen-responsive transcription factors with known edges from the prior in gray and new edges in red (C) The number of interactions for each TF; interaction edges present in the prior that are not in the final network are included in black. The nitrogen TFs knocked out in this work are labeled in blue, and TFs with gene ontology annotations for mitotic cell cycle are annotated in green (D) Gene ontology classification of network interactions by the GO slim biological process terms annotated for the target gene and the regulatory TF (the GO term transcription from RNA pol II is omitted from the annotations for regulatory TFs).

for each regulator-target interaction was calculated. GRNs learned for each condition were combined by rank summing condition-specific confidence scores to create a global confidence score for each potential interaction. All potential interactions are ranked by global confidence score, and a global GRN is constructed from interactions that meet the precision threshold of 0.5, as measured by recovery of known interactions (Figure 6A, Supplemental Data 2). The resulting GRN comprises 6,009 new interactions and 6,009 interactions present in the priors, resulting in a total of 12,018 regulator-target interactions. We find that 3,960 interactions from the priors are not recovered (recall of 0.603). The global GRN comprises an identified regulator for approximately half of all known genes (Supplemental Figure 6A). We find a positive correlation is observed between expression level for a gene and number of regulators for that gene (Supplemental Figure 6B) and that 90% of the identified interactions are activating and 10% are repressing (Supplemental Figure 6C).

Within the nitrogen-regulated TF subnetwork comprising the 11 deleted TFs (Figure 6B) we identify 518 regulator-target interactions, of which 178 are novel, and 340 are present in the priors. This subnetwork contains many features consistent with expectations including co-regulation of targets by the NCR TFs, and redundancy between the SPS-pathway TFs *STP1* and *STP2*. Consistent with the role of TORC1 in regulating nitrogen TFs, the expression of some target genes is altered in rapamycin-treated culture changes as a result of deleting nitrogen TFs (Supplemental Figure 6D). Not all target genes have altered expression when a regulatory TF is deleted in this condition, which implies the presence of redundant regulation. Alternatively, *DUR3* and *TRM5* have altered expression when *RTG1* or *RTG3* are deleted but are not identified as *RTG1* or *RTG3* targets, suggesting that these changes in expression are indirect, and demonstrating the inadequacy of using expression changes to determine regulatory relationships.

Overall, the global GRN has the largest number of target genes for general TFs (*ABF1*, *RAP1*), but we also define regulatory relationships for a total of 148 of the predicted 207 yeast TFs (Figure 6C). The poorest recovery of prior data is found for TFs that regulate environmental responses not included in our experimental design, such as the stress response TF *MSN2* and the mating TF *STE12*, highlighting the necessity of exploration of condition space for complete network reconstruction. Regulators and target genes can be mapped to Gene Ontology (GO) biological process slim terms, which are broad categorizations that facilitate pathway analysis. Ordering GO slim terms by the number of interactions in the learned GRN, we find that for target genes nine of the top ten GO slim terms are metabolism-related (Figure 6D i); in contrast, for regulatory TFs, four of the top ten GO slim terms are DNA replication/repair and cell cycle related (Figure 6D ii).

Identification of Coregulation by Cell Cycle and Environmental Response TFs

Analysis of single cell expression in asynchronous cultures allows detection of cell cycle regulated relationships. The learned global GRN contains 165 genes that are regulated both by nitrogen TFs and by cell cycle TFs (Figure 7A). Many of these regulatory connections are novel;

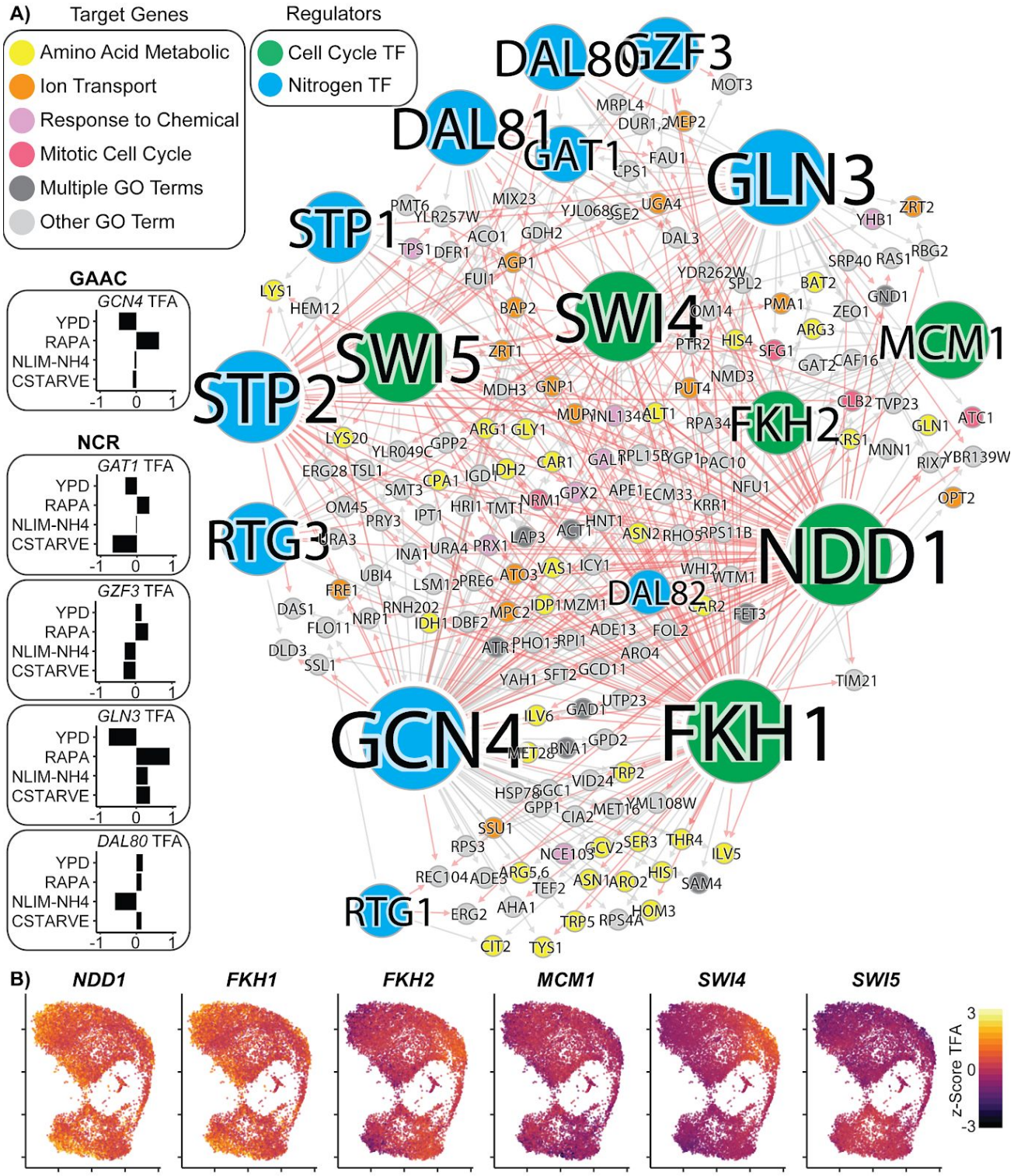


Figure 7: Coordinated regulation of Nitrogen Response and Cell Cycle (A) A gene regulatory network showing target genes that are regulated by at least one nitrogen TF (blue) and at least one cell cycle TF (green). Target gene nodes are colored by GO slim term. Newly inferred regulatory edges are red and known regulatory edges from the prior are in gray. Transcription factor activity (TFA) is calculated from the learned network and then scaled to a z-score over all cells which do not have that TF deleted (e.g. *gcn4Δ* cells are omitted from the calculation for *GCN4* TFA). The mean TFA z-score for four selected conditions is inset for GAAC and NCR TFs (B) TFA for cell cycle TFs for each cell in the YPD growth condition.

likely due to the fact that identifying interactions between metabolism and cell cycle are challenging in asynchronous cultures without single-cell techniques. Of these genes, 37 are annotated with the amino acid metabolic biological process GO term and 21 are annotated with the ion or transmembrane transport biological process GO term. Only 5 are annotated with the mitotic cell cycle biological process GO term, indicating that the majority of the interconnection between cell cycle and nitrogen response genes is due to regulation of metabolism-related genes by cell cycle TFs.

We estimated the TFA for every TF in each cell, using the learned GRN and the single-cell expression matrix. The TFA of nitrogen responsive TFs is principally linked to growth condition as these TFs vary in activity between conditions (Figure 7A), but are generally similar within condition (Supplemental Figure 7A). As expected, we find that cells grown in rich media (YPD) have low TFA for the NCR TFs *GLN3* and *GAT1*, and the GAAC TF *GCN4*. The TFA for these TFs increases substantially upon treatment with rapamycin. The NCR TFs *GZF3* and *DAL80* are not changed with rapamycin treatment, but instead appear to be more responsive to nutrient limitation. By contrast, the estimated TFAs of cell cycle TFs varies within condition (Figure 7B); and are concordant with cell cycle responsive gene expression (Supplemental Figure 7B-D).

Discussion

A Robust scRNAseq and Transcriptional Barcoding Method in Yeast

Since the inception of single-cell RNA sequencing (Tang et al.,), technological advances have resulted in the scale of datasets increasing from tens of cells to tens of thousands in a diversity of organisms. However, scRNAseq in yeast has been comparatively limited in scale in studies published to date (Gasch et al., 2017; Nadal-Ribelles et al., 2019). We present here the first report of large-scale scRNAseq in this widely used model eukaryotic cell. Using a diverse library of transcriptionally barcoded gene deletion strains we were able to efficiently analyze the gene expression state of 38,255 cells using 11 experiments. In addition to facilitating multiplexed analysis of genotypes, transcriptional barcoding provides a facile means of identifying doublet cells within droplets thereby increasing the accuracy of single cell analysis. We envisage that transcriptional barcoding can be broadly implemented to facilitate multiplexed scRNAseq analysis of diverse engineered, and natural, genotypes in yeast, and likely other organisms.

Consistent with our understanding of global gene expression variation first characterized in foundational studies of the transcriptome (DeRisi et al., 1997; Gasch et al., 2000), we find that environmental condition is the primary determinant of the gene expression state of individual yeast cells. However, we observe significant heterogeneity in individual cell gene expression within conditions. Much of this variation can be explained by the mitotic cell-cycle. It is important to note that we do not remove or suppress this cell-cycle driven variance. The cell cycle is itself driven by transcriptional regulators, and our goal is to build a network that integrates cell-cycle regulation with regulated responses to the environment. The ability to access the crosstalk between signalling pathways and the cell cycle program is a key advantage to performing

single-cell sequencing in asynchronous cultures, which bypasses many of the limitations of synchronized bulk sequencing experiments.

Deletion of specific transcription factors results in changes in single cell gene expression for some TFs in some conditions. However, genotypic effects are comparatively minor. We believe that this is due to multiple factors including functional redundancy between TFs, physiological adaptation to the genetic perturbation and the conditional specificity of TFs. It is likely that perturbations that are transiently induced, and result in increased TF activity (McIsaac et al., 2013) may be effective in eliciting detectable responses in gene expression, facilitating causal inference. The use of precise gene deletions does provide several advantages over the use of CRISPR/Cas9-based perturbations as engineered deletions are unambiguous whereas the efficiency of perturbation by CRISPR/Cas9 varies for different guide RNAs.

A Generalizable Framework for GRN Construction Using scRNAseq

Constructing GRNs from single cell gene expression data is a universal goal in all organisms. A yeast single-cell expression matrix has several beneficial properties for design and testing of gene regulatory network inference models as there exist high quality known interactions and TF binding motifs. The issues of data sparsity and low sampling rates are likely to be problems common in experiments in any organism using scRNAseq. We find that techniques that have been developed for normalization and imputation do not improve performance of the additive linear model-based inference of the Inferelator algorithm (Figure 5). However, there are significant opportunities for development of smoothing techniques that would enhance network inference, perhaps targeting latent biophysical parameters like transcription factor activity. It seems reasonable to assume that these biophysical parameters should be stable within the local neighborhood of samples, and the activity calculation that we have used is ill-conditioned and potentially unstable. This is of particular concern when working with undersampled single-cell data and we are actively addressing this issue.

We find that the application of multitask learning is well suited to GRN reconstruction from scRNAseq data. Jointly learning multiple related tasks improves generalization accuracy, especially in scenarios in which datasets are undersampled (Caruana, 1998), and has the desirable side benefit of mitigating the need for complex batch-correction techniques that aim to address technical variation between experiments. Removing batch-effect technical noise from data without suppressing interbatch biological variability remains an unsolved problem, and therefore application of multitask learning approaches to network inference from single-cell data is likely to be generally applicable to integrating scRNAseq data from different cell types and conditions.

A Global GRN for Budding Yeast

Using our scRNAseq dataset, we reconstructed a global GRN with several novel regulatory relationships. Among the most novel of these are cell-cycle associated TFs and targets,

including for TFs like *NDD1* for which prior information is limited. We also identified a number of new regulatory relationships between nitrogen TFs and target genes. Incorporation of additional information into the network inference process, including information about interactions between transcription factors such as functional redundancy and heterodimerization, would likely improve learning of the network. We note that several TFs have few learned targets reflecting the requirement for surveying conditions in which particular TFs are active. For example, *STE12* and *TEC1* are mating-related TFs that we expect to be entirely inactive in our diploid cells; *MSN2* and *YAP1* are stress-responsive TFs that respond to specific stimuli that were not tested in our study. Targeted analysis of the GRN with rationally designed genetic perturbations and environmental conditions will maximize the additional information that can be recovered from futures experiments

Conclusion

Single-cell sequencing is a transformative method for systems biology. To date, scRNAseq has been widely applied to the problem of defining different cell types. However, the ability to simultaneously study the expression of hundreds of genotypes in different conditions, and sample the expression state of thousands of cells, presents a rich source of information for the purpose of GRN reconstruction. Our study implements this approach in budding yeast, the workhorse of systems biology, and establishes a generalizable framework for GRN reconstruction from scRNAseq data in any organism.

Acknowledgements

We would like to thank past and present members of the Gresham and Bonneau labs, as well as Christine Vogel's lab, for discussions and feedback. We thank Tara Rock, Olivia Micci-Smith, and Hana Husic from the NYU Genomics Core facility for troubleshooting suggestions and DNA sequencing services. RB acknowledges support from the Flatiron Institute, the Simons Foundation, and the US NIH. DG is funded by the NIH (R01GM107466) and NSF (MCB1818234).

Author Contributions

Conceptualization: CJ, DG, and RB. Formal analysis: CJ, DC, and GS. Funding acquisition: DG and RB. Investigation: CJ and DG. Methodology: CJ, DC, and GS. Software: CJ, DC, and GS. Supervision: DG and RB. Writing – original draft: CJ and DC. Writing – review & editing: CJ, DC, GS, DG, and RB

Declaration of Interests

The authors declare no competing interests.

References

Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, Pak RA, Gray AN, Gross CA, Dixit A, Parnas O, Regev A, Weissman JS. 2016. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**:1867–1882.e21.

Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine J-C, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**:1083–1086.

Airoldi EM, Miller D, Athanasiadou R, Brandt N, Abdul-Rahman F, Neymotin B, Hashimoto T, Bahmani T, Gresham D. 2016. Steady-state and dynamic gene expression programs in Saccharomyces cerevisiae in response to variation in environmental nitrogen. *Mol Biol Cell* **27**:1383–1396.

Andréasson C, Ljungdahl PO. 2002. Receptor-mediated endoproteolytic activation of two transcription factors in yeast. *Genes Dev* **16**:3158–3172.

Arrieta-Ortiz ML, Hafemeister C, Bate AR, Chu T, Greenfield A, Shuster B, Barry SN, Gallitto M, Liu B, Kacmarczyk T, Santoriello F, Chen J, Rodrigues CDA, Sato T, Rudner DZ, Driks A, Bonneau R, Eichenberger P. 2015. An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network. *Mol Syst Biol* **11**:839.

Athanasiadou R, Neymotin B, Brandt N, Wang W, Christiaen L, Gresham D, Tranchina D. 2019. A complete statistical model for calibration of RNA-seq counts using external spike-ins and maximum likelihood theory. *PLoS Comput Biol* **15**:e1006794.

Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, Nainys J, Wu K, Kiseliovas V, Setty M, Choi K, Fromme RM, Dao P, McKenney PT, Wasti RC, Kadaveru K, Mazutis L, Rudensky AY, Pe'er D. 2018. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* **174**:1293–1308.e36.

Barabási A-L, Gulbahce N, Loscalzo J. 2011. Network medicine: a network-based approach to human disease. *Nat Rev Genet* **12**:56–68.

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. *J Stat Mech* **2008**:P10008.

Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, Thorsson V. 2006. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol* **7**:R36.

Brennecke P, Anders S, Kim JK, Kołodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC, Heisler MG. 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods* **10**:1093–1095.

Carmona-Gutierrez D, Ruckenstuhl C, Bauer MA, Eisenberg T, Büttner S, Madeo F. 2010. Cell death in yeast: growing applications of a dying buddy. *Cell Death Differ* **17**:733–734.

Caruana R. 1998. Multitask Learning In: Thrun S, Pratt L, editors. Learning to Learn. Boston, MA: Springer US. pp. 95–133.

Castro DM, de Veaux NR, Miraldi ER, Bonneau R. 2019. Multi-study inference of regulatory networks for more accurate models of gene regulation. *PLoS Comput Biol* **15**:e1006591.

Chan TE, Stumpf MPH, Babtie AC. 2017. Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Syst* **5**:251–267.e3.

Chen M, Zhou X. 2018. VIPER: variability-preserving imputation for accurate gene expression

recovery in single-cell RNA sequencing studies. *Genome Biol* **19**:196.
Chen S, Mar JC. 2018. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics* **19**:232.
Ciofani M, Madar A, Galan C, Sellars M, Mace K, Pauli F, Agarwal A, Huang W, Parkurst CN, Muratet M, Newberry KM, Meadows S, Greenfield A, Yang Y, Jain P, Kirigin FK, Birchmeier C, Wagner EF, Murphy KM, Myers RM, Bonneau R, Littman DR. 2012. A Validated Regulatory Network for Th17 Cell Specification. *Cell* **151**:289–303.
Colman-Lerner A, Chin TE, Brent R. 2001. Yeast Cbk1 and Mob2 activate daughter-specific genetic programs to induce asymmetric cell fates. *Cell* **107**:739–750.
Cox KH, Rai R, Distler M, Daugherty JR, Coffman JA, Cooper TG. 2000. Saccharomyces cerevisiae GATA sequences function as TATA elements during nitrogen catabolite repression and when Gln3p is excluded from the nucleus by overproduction of Ure2p. *J Biol Chem* **275**:17611–17618.
Davidson EH. 2012. Gene Activity in Early Development. Elsevier.
de Boer CG, Hughes TR. 2012. YeTFaSCo: a database of evaluated yeast transcription factor sequence specificities. *Nucleic Acids Res* **40**:D169–79.
DeRisi JL, Iyer VR, Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**:680–686.
Didion T, Regenberg B, Jørgensen MU, Kielland-Brandt MC, Andersen HA. 1998. The permease homologue Ssy1p controls the expression of amino acid and peptide transporter genes in Saccharomyces cerevisiae. *Mol Microbiol* **27**:643–650.
Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, Adamson B, Norman TM, Lander ES, Weissman JS, Friedman N, Regev A. 2016. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**:1853–1866.e17.
Eriksson PR, Ganguli D, Nagarajavel V, Clark DJ. 2012. Regulation of histone gene expression in budding yeast. *Genetics* **191**:7–20.
Filtz TM, Vogel WK, Leid M. 2014. Regulation of transcription factor activity by interconnected post-translational modifications. *Trends Pharmacol Sci* **35**:76–85.
Fu Y, Jarboe LR, Dickerson JA. 2011. Reconstructing genome-wide regulatory network of E. coli using transcriptome data and predicted transcription factor activities. *BMC Bioinformatics* **12**:233.
Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Molecular Biology of the Cell*. doi:10.1091/mbc.11.12.4241
Gasch AP, Yu FB, Hose J, Escalante LE, Place M, Bacher R, Kanbar J, Ciobanu D, Sandor L, Grigoriev IV, Kendziorski C, Quake SR, McClean MN. 2017. Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. *PLoS Biol* **15**:e2004050.
Georis I, Feller A, Vierendeels F, Dubois E. 2009. The Yeast GATA Factor Gat1 Occupies a Central Position in Nitrogen Catabolite Repression-Sensitive Gene Activation. *Mol Cell Biol* **29**:3803–3815.
Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, Arkin AP, Astromoff A, Bakkoury ME, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian K-D, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Güldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kötter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M,

Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang C-Y, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M. 2002. Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**:387.

Godard P, Urrestarazu A, Vissers S, Kontos K, Bontempi G, van Helden J, André B. 2007. Effect of 21 Different Nitrogen Sources on Global Gene Expression in the Yeast Saccharomyces cerevisiae. *Mol Cell Biol* **27**:3065–3086.

González A, Hall MN. 2017. Nutrient sensing and TOR signaling in yeast and mammals. *EMBO J* **36**:397–408.

Gray JV, Petsko GA, Johnston GC, Ringe D, Singer RA, Werner-Washburne M. 2004. “Sleeping beauty”: quiescence in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **68**:187–206.

Greenfield A, Hafemeister C, Bonneau R. 2013. Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics* **29**:1060–1067.

Greenfield A, Madar A, Ostrer H, Bonneau R. 2010. DREAM4: Combining genetic and dynamic information to identify biological networks and dynamical models. *PLoS One* **5**:e13397.

Grün D, Kester L, van Oudenaarden A. 2014. Validation of noise models for single-cell transcriptomics. *Nat Methods* **11**:637–640.

Hicks SC, Townes FW, Teng M, Irizarry RA. 2018. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* **19**:562–578.

Hinnebusch AG. 2005. Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* **59**:407–450.

Hofman-Bang J. 1999. Nitrogen catabolite repression in Saccharomyces cerevisiae. *Mol Biotechnol* **12**:35–73.

Hu JX, Thomas CE, Brunak S. 2016. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* **17**:615–629.

Hwang B, Lee JH, Bang D. 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* **50**:96.

Iraqui I, Vissers S, Bernard F, de Craene JO, Boles E, Urrestarazu A, André B. 1999. Amino acid signaling in Saccharomyces cerevisiae: a permease-like sensor of external amino acids and F-Box protein Grr1p are required for transcriptional induction of the AGP1 gene, which encodes a broad-specificity amino acid permease. *Mol Cell Biol* **19**:989–1001.

Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, Amit I. 2016. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**:1883–1896.e15.

Jia Y, Rothermel B, Thornton J, Butow RA. 1997. A basic helix-loop-helix-leucine zipper transcription complex in yeast functions in a signaling pathway from mitochondria to the nucleus. *Mol Cell Biol* **17**:1110–1117.

Johnston GC, Pringle JR, Hartwell LH. 1977. Coordination of growth with cell division in the yeast Saccharomyces cerevisiae. *Exp Cell Res* **105**:79–98.

Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, Taipale J. 2011. Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* **9**:72–74.

Komeili A, Wedaman KP, O’Shea EK, Powers T. 2000. Mechanism of metabolic control. Target of rapamycin signaling links nitrogen quality to the activity of the Rtg1 and Rtg3 transcription factors. *J Cell Biol* **151**:863–878.

Lam KY, Westrick ZM, Müller CL, Christiaen L, Bonneau R. 2016. Fused Regression for Multi-source Gene Regulatory Network Inference. *PLoS Comput Biol* **12**:e1005157.

Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* **11**:733–739.
Liao X, Butow RA. 1993. RTG1 and RTG2: two yeast genes required for a novel path of communication from mitochondria to the nucleus. *Cell* **72**:61–71.
Li WV, Li JJ. 2018. An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nat Commun* **9**:997.
Ljungdahl PO. 2009. Amino-acid-induced signalling via the SPS-sensing pathway in yeast. *Biochem Soc Trans* **37**:242–247.
Loewith R, Hall MN. 2011. Target of Rapamycin (TOR) in Nutrient Signaling and Growth Control. *Genetics* **189**:1177–1201.
Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**:1202–1214.
Madar A, Greenfield A, Vanden-Eijnden E, Bonneau R. 2010. DREAM3: Network Inference Using Dynamic Context Likelihood of Relatedness and the Inferelator. *PLoS One* **5**:e9803.
Ma S, Kemmeren P, Gresham D, Statnikov A. 2014. De-Novo Learning of Genome-Scale Regulatory Networks in S. cerevisiae. *PLoS One* **9**:e106479.
McIsaac RS, Scott McIsaac R, Oakes BL, Wang X, Dummit KA, Botstein D, Noyes MB. 2013. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Research*. doi:10.1093/nar/gks1313
Milias-Argeitis A, Oliveira AP, Gerosa L, Falter L, Sauer U, Lygeros J. 2016. Elucidation of Genetic Interactions in the Yeast GATA-Factor Network Using Bayesian Model Selection. *PLoS Comput Biol* **12**:e1004784.
Miller D, Brandt N, Gresham D. 2018. Systematic identification of factors mediating accelerated mRNA degradation in response to changes in environmental nitrogen. *PLoS Genet* **14**:e1007406.
Miraldi ER, Pokrovskii M, Watters A, Castro DM, De Veaux N, Hall JA, Lee J-Y, Ciofani M, Madar A, Carriero N, Littman DR, Bonneau R. 2019. Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells. *Genome Res*. doi:10.1101/gr.238253.118
Mittal N, Guimaraes JC, Gross T, Schmidt A, Vina-Vilaseca A, Nedialkova DD, Aeschimann F, Leidel SA, Spang A, Zavolan M. 2017. The Gcn4 transcription factor reduces protein synthesis capacity and extends yeast lifespan. *Nat Commun* **8**:457.
Mueller PP, Hinnebusch AG. 1986. Multiple upstream AUG codons mediate translational control of GCN4. *Cell* **45**:201–207.
Nadal-Ribelles M, Islam S, Wei W, Latorre P, Nguyen M, de Nadal E, Posas F, Steinmetz LM. 2019. Sensitive high-throughput single-cell RNA-seq reveals within-clonal transcript correlations in yeast populations. *Nature Microbiology*. doi:10.1038/s41564-018-0346-9
Nagarajan S, Kruckeberg AL, Schmidt KH, Kroll E, Hamilton M, McInnerney K, Summers R, Taylor T, Rosenzweig F. 2014. Uncoupling reproduction from metabolism extends chronological lifespan in yeast. *Proc Natl Acad Sci U S A* **111**:E1538–47.
Natarajan K, Meyer MR, Jackson BM, Slade D, Roberts C, Hinnebusch AG, Marton MJ. 2001. Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol Cell Biol* **21**:4347–4368.
Rødkær SV, Færgeman NJ. 2014. Glucose- and nitrogen sensing and regulatory mechanisms in Saccharomyces cerevisiae. *FEMS Yeast Res* **14**:683–696.

Ruiz-Roig C, Noriega N, Duch A, Posas F, de Nadal E. 2012. The Hog1 SAPK controls the Rtg1/Rtg3 transcriptional complex activity by multiple regulatory mechanisms. *Mol Biol Cell* **23**:4286–4296.

Siahpirani AF, Roy S. 2017. A prior-based integrative framework for functional transcriptional regulatory network inference. *Nucleic Acids Res* **45**:e21.

Silverman SJ, Petti AA, Slavov N, Parsons L, Briehof R, Thiberge SY, Zenklusen D, Gandhi SJ, Larson DR, Singer RH, Botstein D. 2010. Metabolic cycling in single yeast cells from unsynchronized steady-state populations limited on glucose or phosphate. *Proc Natl Acad Sci U S A* **107**:6946–6951.

Soneson C, Robinson MD. 2018. Bias, robustness and scalability in single-cell differential expression analysis. *Nat Methods* **15**:255–261.

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell* **9**:3273–3297.

Stanbrough M, Magasanik B. 1995. Transcriptional and posttranslational regulation of the general amino acid permease of Saccharomyces cerevisiae. *J Bacteriol* **177**:94–102.

Talarek N, Gueydon E, Schwob E. 2017. Homeostatic control of START through negative feedback between Cln3-Cdk1 and Rim15/Greatwall kinase in budding yeast. *Elife* **6**. doi:10.7554/eLife.26233

Tchourine K, Vogel C, Bonneau R. 2018. Condition-Specific Modeling of Biophysical Parameters Advances Inference of Regulatory Networks. *Cell Rep* **23**:376–388.

Teixeira MC, Monteiro PT, Palma M, Costa C, Godinho CP, Pais P, Cavalheiro M, Antunes M, Lemos A, Pedreira T, Sá-Correia I. 2018. YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in Saccharomyces cerevisiae. *Nucleic Acids Res* **46**:D348–D353.

Tu BP, Kudlicki A, Rowicka M, McKnight SL. 2005. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science* **310**:1152–1158.

van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, Burdziak C, Moon KR, Chaffer CL, Pattabiraman D, Bierie B, Mazutis L, Wolf G, Krishnaswamy S, Pe'er D. 2018. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**:716–729.e27.

Ward LD, Bussemaker HJ. 2008. Predicting functional transcription factor binding through alignment-free and affinity-based analysis of orthologous promoter sequences. *Bioinformatics* **24**:i165–71.

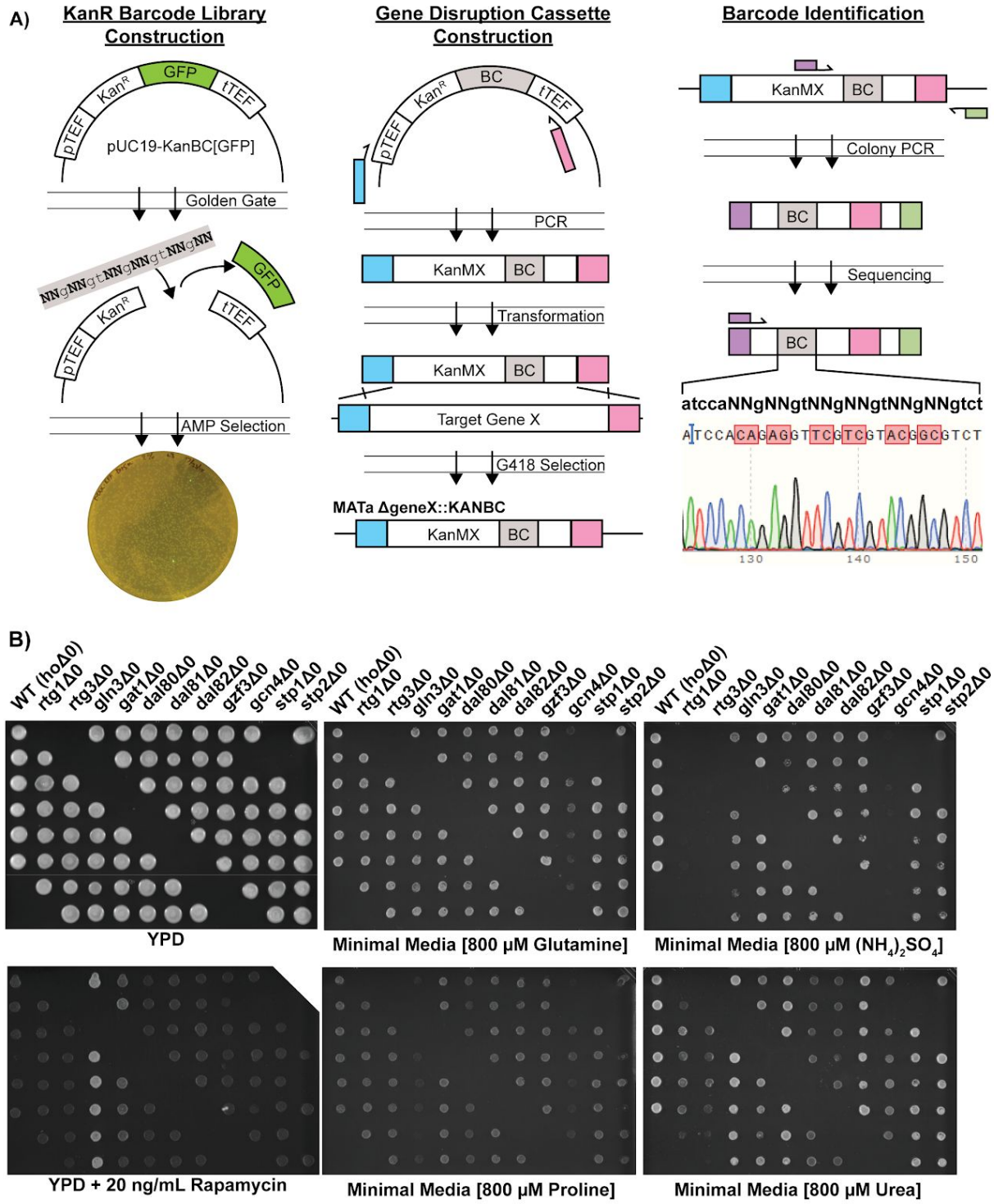
Wilkins O, Hafemeister C, Plessis A, Holloway-Phillips M-M, Pham GM, Nicotra AB, Gregorio GB, Jagadish SVK, Septiningsih EM, Bonneau R, Purugganan M. 2016. EGRINs (Environmental Gene Regulatory Influence Networks) in Rice That Function in the Response to Water Deficit, High Temperature, and Agricultural Environments. *Plant Cell* **28**:2365–2384.

Xu C, Su Z. 2015. Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics* **31**:1974–1980.

Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**:14049.

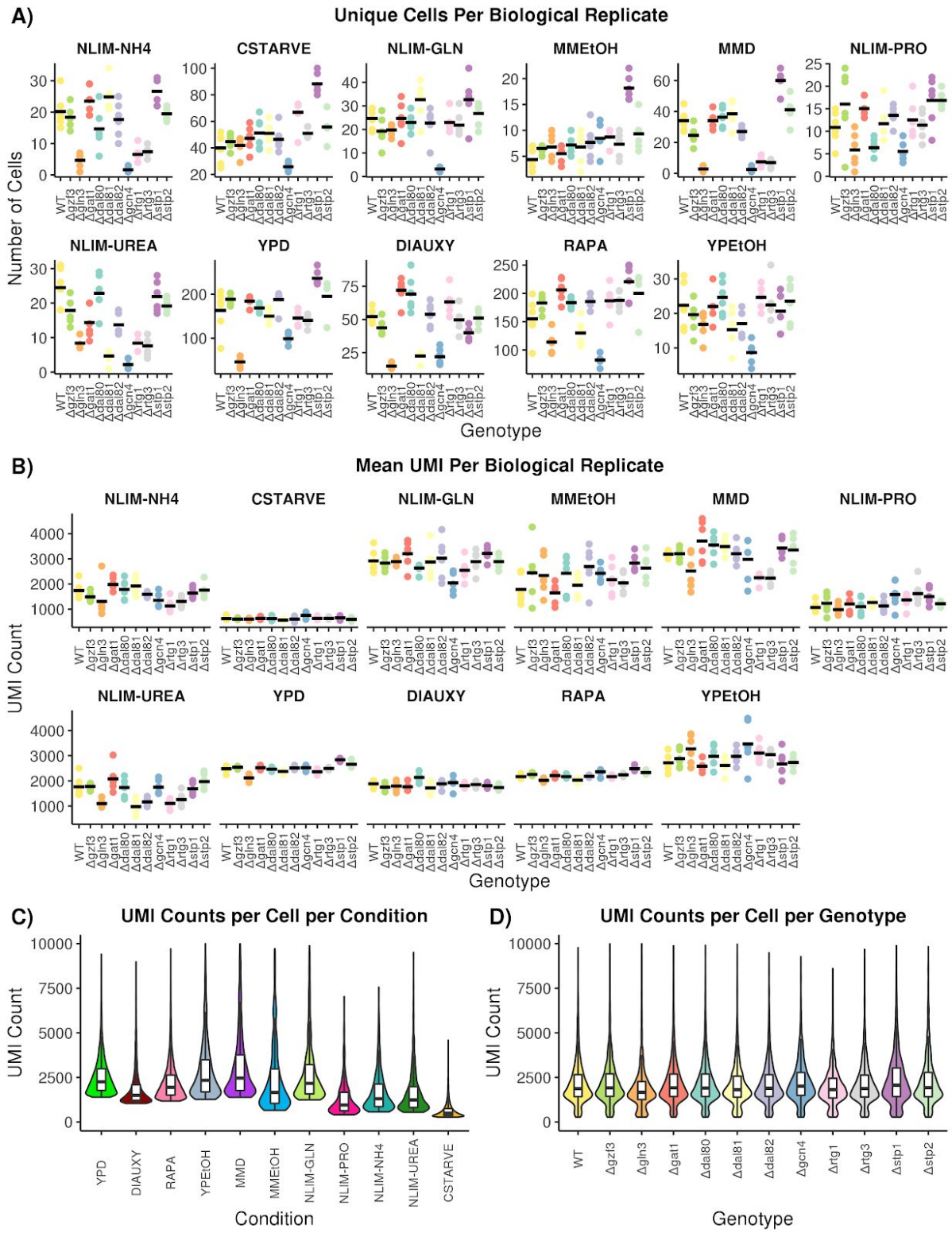
Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, Mazutis L. 2017. Single-cell barcoding and sequencing using droplet microfluidics. *Nat Protoc* **12**:44.

Zinzalla V, Graziola M, Mastriani A, Vanoni M, Alberghina L. 2007. Rapamycin-mediated G1 arrest involves regulation of the Cdk inhibitor Sic1 in Saccharomyces cerevisiae. *Mol Microbiol* **63**:1482–1494.



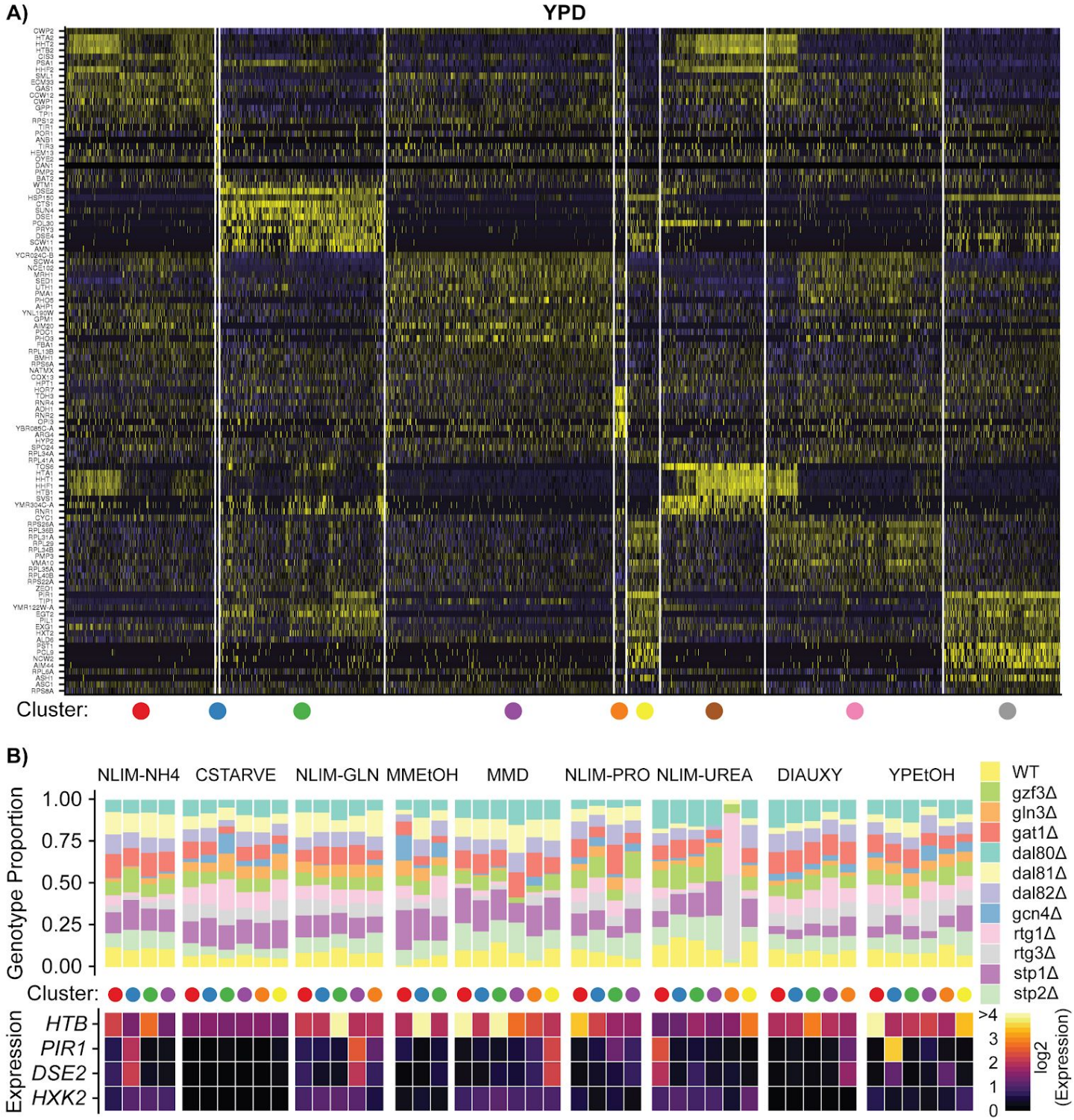
Supplemental Figure 1: Strain Construction Workflow and Validation

(A) Construction of KanMX[BC] deletion cassettes containing degenerate barcodes in the 3' untranslated region, followed by transformation into yeast to create a transcription factor knockout array (B) Spotting of the deletion array onto YPD (control) and media containing different nitrogen sources plates. Each column is a separate TF deletion, and each spot is a uniquely barcoded biological replicate. There are 6 biological replicates for each of 12 genotypes, for a total of 72 unique strains in the array.

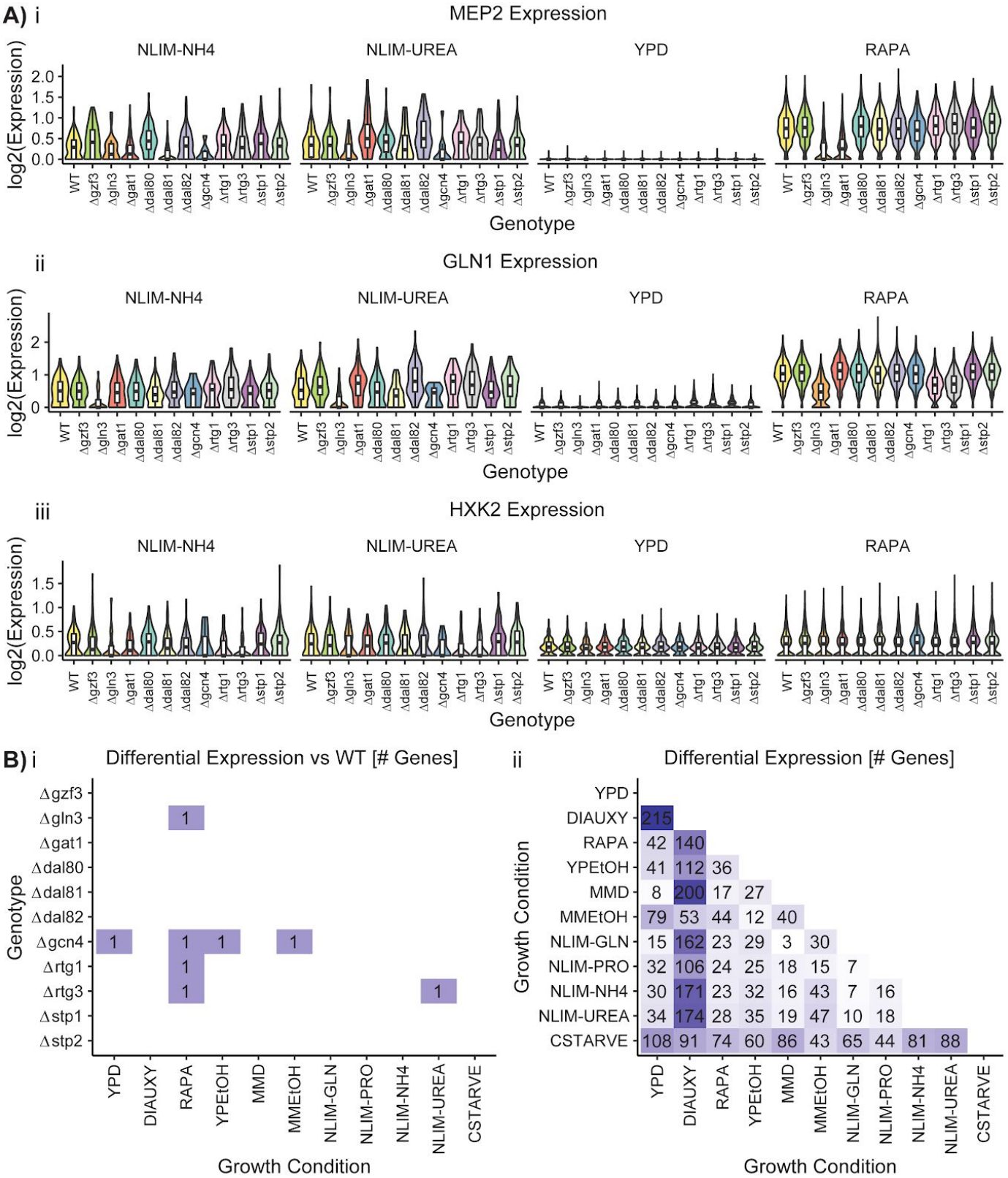


Supplemental Figure 2: Quality Control of Single-Cell RNA Sequencing Data

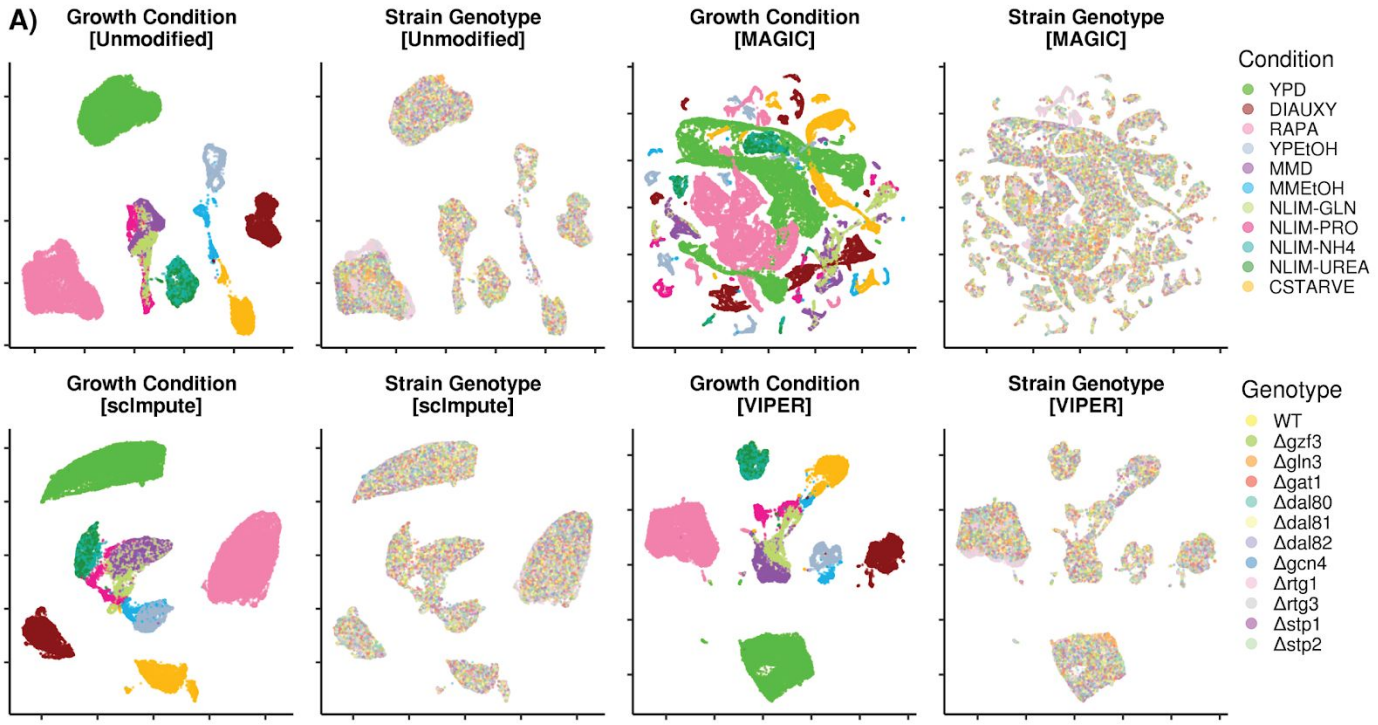
(A) The number of cells that pass all quality control and preprocessing filters for each growth condition. Each genotype has multiple independently barcoded biological replicates that are plotted separately within each condition. The mean number of cells for each genotype within a condition is plotted as a horizontal line. (B) The mean count of unique transcripts, determined by UMI, for each growth condition. Biological replicates are plotted separately for each genotype within each condition. The mean number of transcripts per genotype is plotted as a horizontal line. (C) The distribution of unique transcript count per cell across all growth conditions. (D) The distribution of unique transcript counts per cell across all genotypes.



Supplemental Figure 3 (A) Gene expression heatmap of genes that are specific for clusters of cells in YPD (B) Summary of clustered single cells within growth conditions (i) Proportion of cells within each cluster that consist of a specific strain genotype (ii) The mean within each cluster of log-transformed, normalized expression of the G1- and S-phase marker genes, as well as a hexokinase gene *HXK2* that is unlikely to be responding to cell cycle

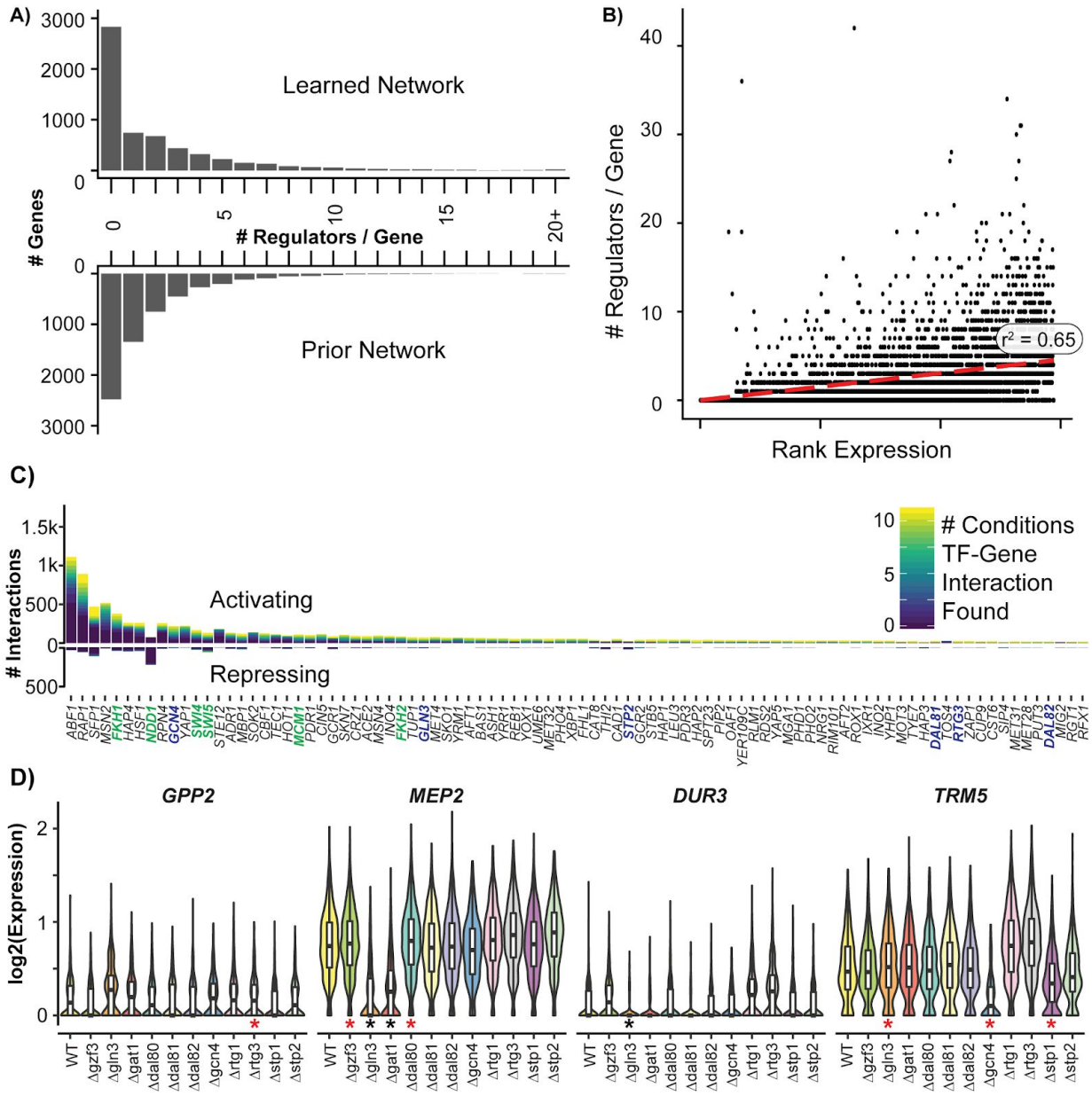


Supplemental Figure 4: (A) Distribution by strain genotype of the $\log_2$ batch-normalized expression of the ammonium permease gene *MEP2,* the glutamine synthetase gene *GLN1*, and the hexokinase gene *HKX2* in YPD, RAPA, and the nitrogen limitation growth conditions. (B) Number of differentially expressed genes identified using pairwise Welch t-tests (fold change > 1.5; p.adj < 0.05) between (i) WT and TF deletion strains or between (ii) growth conditions



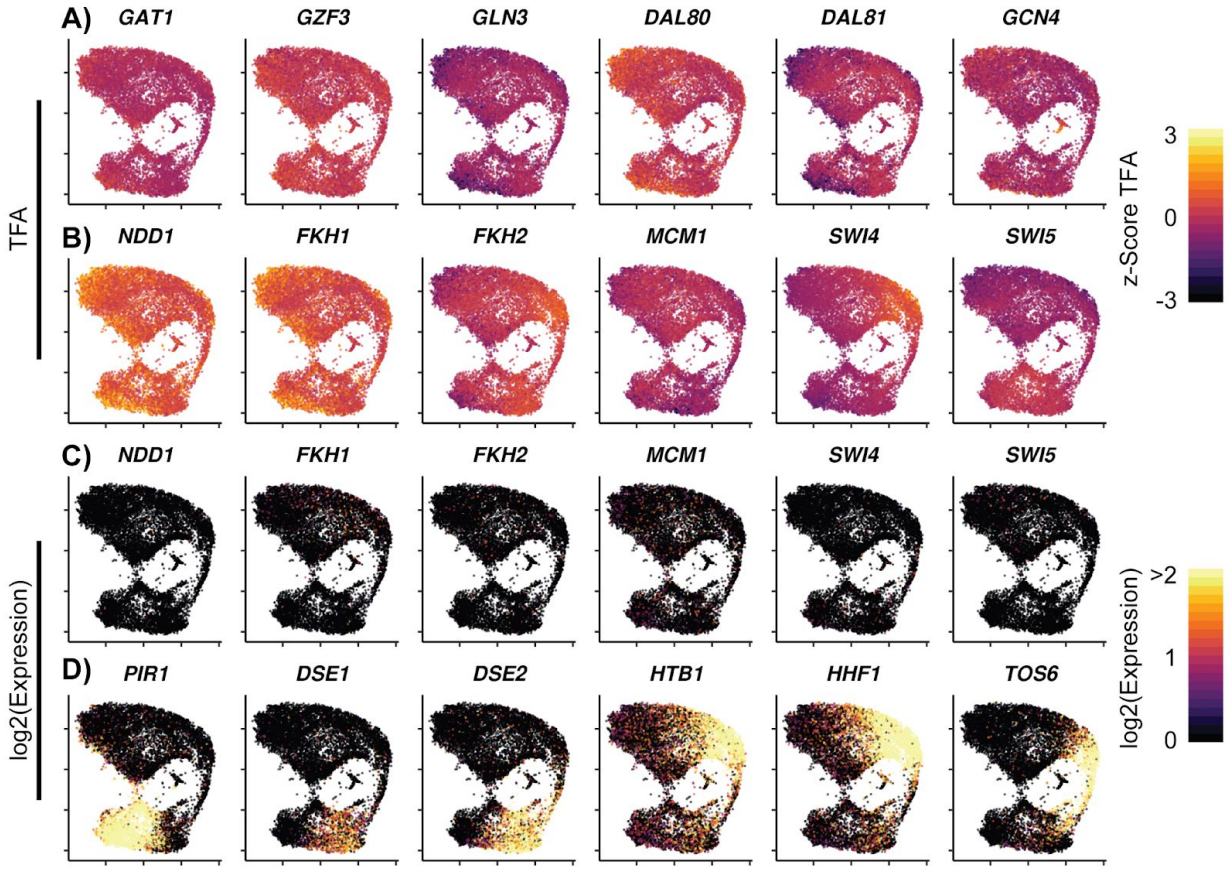
Supplemental Figure 5: Low-Dimensional Clustering of Imputed Data

(A) Scatter plot after UMAP into 2-dimensional space. Unmodified data is is compared to imputation methods to recover missing data using ScImpute, MAGIC, and VIPER.



Supplemental Figure 6. Summary of Learned GRN

(A) Histogram of the number of regulators per target gene in the learned and prior network (B) Rank correlation plot between the mean expression of a gene and the number of regulators for that gene; r^2 is calculated using Spearman's Rank correlation (C) The number of interactions for a TF that are activating (positive) or repressing (negative). Bars are colored by the number of separate conditions in which the interaction is identified, with condition-specific networks selected using a precision threshold of 0.5 (D) Distribution by strain genotype of the $\log_2$ batch-normalized expression following rapamycin treatment of the glycerol phosphatase gene *GPP2*, the ammonium permease gene *MEP2*, the urea transporter gene *DUR3*, and the tRNA methyltransferase gene *TRM5*. TF deletion strains corresponding to regulatory TFs from the GRN are identified with asterisks; red asterisks indicate newly learned interactions and black asterisks indicate interactions known from the prior.



Supplemental Figure 7. Cell Cycle TF Activity Clusters within Growth Conditions

Cells grown in YPD are plotted after UMAP with (A-B) z-Score of the calculated transcription factor activity (TFA) based on the learned network for (A) select nitrogen TFs and (B) cell cycle TFs (C) Log_2 of the expression of cell cycle TFs (D) Log_2 of the expression of cell cycle target genes

METHODS

Requests for strains and reagents should be directed to David Gresham (dgresham@nyu.edu). Requests related to computational analysis and code should be directed to Richard Bonneau (rb133@nyu.edu). There are no restrictions on the materials or the code used in this work. All materials are released under CC-BY 4.0 and all code is available under the permissive MIT or BSD licenses.

YEAST STRAIN CONSTRUCTION AND GROWTH

All yeast strains were generated from the prototrophic FY4 (MAT**a**) or FY5 (MAT**α**) background strains. Yeast were transformed using the standard lithium acetate transformation protocol (Gietz and Schiestl, 2007). *E. coli* were transformed using the standard chemically competent transformation protocol. Plasmid constructions were confirmed by sanger sequencing. Yeast genotypes, plasmid sequences, and oligonucleotide sequences are provided as supplemental tables 1-3. Media formulations are provided as supplemental table 4.

Construction of Barcoded Deletion Cassettes

The deletion cassette plasmid was constructed by amplifying pTEF::KAN^R from pUG6 (Euroscarf) and tTEF from pUG6, with an overlapping junction between KAN^R and tTEF containing two BbsI sites for golden-gate mediated barcode cloning. These pieces were assembled into pUC19 using gibson isothermal assembly to generate DGP304. This plasmid was then modified by linearizing with BamHI and XbaI, amplifying a bacterial GFP expression cassette from pWS158 (Addgene), and assembled using gibson isothermal assembly to generate DGP306.

Gene deletion barcodes were created by synthesizing an oligonucleotide containing flanking PCR handles (M13F and M13R), flanking BbsI sites for golden gate cloning, and the degenerate sequence caNNgNNgtNNgNNgtNNgNNgt. The mixture of oligonucleotides was double-stranded using *E. coli* DNA Polymerase I, Large (Klenow) fragment. Klenow buffer (1x NEB Buffer 2.1 [NEB #B7202S]) was mixed with 250 nM barcode oligonucleotide, 250 nM M13R primer, 200 nM/each dNTP [NEB #N0447S], incubated at 80°C and slowly cooled to room temperature. The DNA Polymerase I, Large (Klenow) Fragment (NEB #M0210S) was added to 0.1 U/μL and the reaction was incubated at 37°C for 30 minutes. The polymerase was heat-inactivated by placing the reaction at 75°C for 20 minutes. The resulting dsDNA cassette was used with no further cleanup.

The barcode was inserted into the 3' untranslated region of the pTEF::KAN^R::tTEF yeast selection marker cassette in DGP306 by BbsI-mediated golden gate cloning. A golden gate reaction was prepared with 1x Thermo FastDigest Buffer [Thermo #ER1011], 1 mM ATP, 10 mM DTT, 2 U/μL T4 DNA Ligase [NEB #M0202S], 1 U/μL BpiI [Thermo #ER1011], 10 ng/μL

DGP306, 25 nM barcode dsDNA, and incubated in a thermocycler using the following program: 37°C 20 minutes; 25x cycles of 37°C for 5 minutes and 16°C for 5 minutes; 37°C for 20 minutes; 80°C for 20 minutes. An additional 1 U/μL BpiI was then added to the reaction mix and incubated at 37°C for 30 minutes to linearize any remaining uncloned plasmid.

The golden gate cloning reaction was transformed into One Shot TOP10 *E. coli* (ThermoFisher #C404003). Cloning and transformation efficiency was estimated by plating 2% of the transformation onto LB + ampicillin plate and counting GFP^+ and GFP^- colonies. The remainder of the reaction was inoculated into 200mL molten LB + ampicillin + 0.6% (w/v) SeaPrep Agarose (Lonza 50302) media, thoroughly mixed, snap cooled in an ice bucket, and incubated overnight at 37°C. The soft agar culture was then collected by centrifugation, washed with PBS, and resuspended in 2mL 50% glycerol. 100μL of this mixture was used to inoculate a culture of 100mL LB + ampicillin and the remainder stored at -80°C in aliquots. The 100mL culture was grown for 8 hours at 37°C, harvested, washed with PBS, and stored at -20°C until midiprepped (Qiagen) according to the manufacturer's protocol.

Construction of a Barcoded Transcription Factor Deletion Array

The degenerate barcoded plasmid was used as template for PCR using primers containing gene-specific targeting homology arms (1x NEB Q5 Master Mix #M0494S, 1 ng template plasmid, 250 nM/each oligo). The PCR amplicon was then transformed into FY4 and plated on YPD+G418 to select transformants. Transformants containing the gene deletion were confirmed using colony PCR and gene-specific primers and a KANR primer. PCR products of correct transformants were cleaned using silica spin columns (Qiagen) according to the manufacturer's protocol and the barcode identified by Sanger sequencing. At least 6 uniquely barcoded strains (i.e. biological replicates) were generated for each genotype, with the criteria that each barcode had to differ by at least 3 bases, ensuring that the probability of barcode collisions is extremely low.

The plasmid DGP328 (pTEF::NATR::tTEF) was used as template for PCR using primers containing the same gene-specific targeting homology. The PCR amplicon was transformed into FY5 and plated on YPD+nourseothricin. Positive transformants were confirmed using colony PCR with gene-specific primers and a NATR primer.

FY4-derived MATa strains were arrayed in a 96-well plate (Corning 3788) and then pinned (V & P Scientific #VP407FP12) onto YPD in an OmniTray (Nunc 165218). FY5-derived MATα strains were arrayed in a 96-well plate so that the same gene was disrupted in matching wells of the MATa and MATα plates and then pinned onto YPD. These arrays are grown overnight at 30°C. The MATa array and MATα array were then pinned to the same YPD plate to create spots where MATa and MATα strains were overlaid. The plate layout was designed so that some locations had only MATa strains, only MATα strains, or no strains, to control for mating, contamination, and the efficacy of diploid selection. The mating array was grown overnight at 30°C to allow mating to occur and then pinned to a YPD+G418+nourseothricin plate to select for

MATa/MATα diploids. This diploid selection plate was grown overnight and then pinned to a YPD+G418+nourseothricin plate for a second round of diploid selection. The second diploid selection plate was grown overnight at 30°C and then pinned to a YPD+G418+nourseothricin plate for a third round of diploid selection at 30°C. This plate was then pinned to several replicate 96 well round-bottom plates containing 200 µL YPD+G418+nourseothricin in each well. These plates were cultured with shaking overnight at 30°C, then centrifuged and the media aspirated. The cells were resuspended in 50% glycerol and the plates stored at -80°C.

Culturing

The barcoded deletion array was pinned from a frozen stock plate at -80°C onto a YPD plate for recovery and grown overnight at 30°C. The first recovery plate was then pinned to a second recovery YPD plate and grown overnight at 30°C. The second recovery plate was pinned to a 96 well round-bottom plate containing 200 µL YPD in each well and grown overnight at 30°C. The cultures from this plate were pooled, washed 2x with 50mL PBS, and then resuspended in 1mL PBS. 250µL of the washed cells was used to inoculate 50mL of the relevant media for the specific experimental condition in a shake flask. These flasks were grown for 4 hours. Cell count per mL at this time was determined using a Beckman Coulter Z2 Particle Counter #6605700. A volume of culture containing ~10^9 cells was collected and the cells pelleted by centrifugation. These cells were immediately resuspended in 1mL RNALater (Qiagen #76104), washed 2x with 1mL RNALater and resuspended in a final volume of 500µL RNALater. This suspension was stored at -20°C for 12 to 72 hours. The experiment grown to diauxic shift was grown for 10 hours. We confirmed that glucose in the media was exhausted between hour 9 and hour 10 using a hexokinase-based assay (R-Biopharm #10716251035). All other steps of harvesting cells were identical to the 4 hour experiments.

SINGLE-CELL SEQUENCING AND NETWORK INFERENCE

All steps below used RNAse-free reagents.

Single Cell Library Preparation

Cells stored in RNALater were removed from -20°C and ~10^7 cells were washed 2x with 1mL spheroplasting buffer (50 mM Sodium Phosphate pH 7.5, 1M Sorbitol, 10 mM EDTA, 2 mM DTT, 100 µg/mL BSA). Cells from fermentative phase growth cultures were then resuspended in 100µL spheroplasting buffer + 0.1 U Zymolyase 100T (Zymo Research #E1004). Cells from respiratory phase growth cultures or starvation cultures were resuspended in 100µL spheroplasting buffer + 0.25U Zymolyase 100T. The spheroplasting reaction was incubated at 37°C for exactly 20 minutes, and then the spheroplasted cells were pelleted and resuspended in 500µL RNALater for 5 minutes on ice. After this incubation the spheroplasted cells were pelleted and washed 3x with 1mL wash buffer (10mM TRIS pH 8, 1M Sorbitol, 100µg/mL BSA) and resuspended in 1mL wash buffer. The cells were visualized to confirm spheroplasting and

counted using a hemocytometer. A dilution equal to ~5 x 10^6 cells/mL in wash buffer wa prepared and then immediately used for single cell isolation.

Single cell library preparation was done using the 10x Genomics Chromium 3' v2 Single Cell Gene Expression Kit (10x Genomics #120237), following the kit protocol. 66.2µL of single-cell master mix was prepared to which 27.7µL H_2O was added. The microfluidic Chromium Single Cell A Chip (10x Genomics #120236) wa then prepared for use. 6µL of prepared spheroplast cell suspension was added to the single-cell master mix, and then immediately transferred to the microfluidics chip. Hydrogel beads and partitioning oil were added according to the manufacturer's protocol, and the cells were encapsulated with hydrogel beads using the 10x Genomics Chromium Controller. Following emulsification, reverse transcription and cleanup was performed according to the manufacturer's protocol. Whole transcriptome amplification was performed using a total of 10 cycles of PCR. Cleanup, fragmentation, adapter ligation, and indexing was performed according to the manufacturer's protocol, using 8-10 cycles of PCR for the indexing reaction.

Transcribed barcodes were amplified from the whole transcriptome amplification prior to fragmentation. The KAN^R transcript containing the genotype barcode was amplified in a reaction(1x KAPA HiFi Hotstart Readymix [Kapa #KK2602], 200 nM/each primer, 1 µL 10x whole-transcriptome DNA), using 6 cycles of PCR (98°C for 3:00; 6 cycles of 98°C for 0:20, 63°C for 0:20, and 72°C for 0:20min; 72°C for 1:00min). The amplicon pool was then purified with 1x volume of SPRIselect beads (Beckman Coulter #B23317) and eluted into 24µL H_2O. To this eluate, 25µL of 2x KAPA HiFi Hotstart Readymix was added, as well as 200nM/each indexing primers. The indexing reaction was cycled for 8-10 cycles of PCR, using the 10x Genomics indexing PCR reaction settings (98°C for 0:45; 8-10x cycles of 98°C for 0:20, 54°C for 0:30, and 72°C for 0:20; 72°C for 1:00).

Library fragment sizes were determined using a High Sensitivity D1000 Screentape (Agilent #5067-5584) and quantified with the KAPA illumina library quantification system (Roche #KK4953) on a Roche lightcycler 480. Libraries from each condition were pooled so that 99% of the pool consisted of the single-cell transcriptome library and 1% of the pool consisted of the genotype barcode amplicon. Samples were then pooled for multiplex sequencing on an Illumina NextSeq 500 with the NextSeq 500/550 v2.5 High Output 150 Cycle kit (Illumina #20024907), using the sequencing parameters recommended by 10x Genomics (Read 1: 26bp, Read 2: 98bp, Index 2: 8bp) and standard illumina read and indexing primers.

Processing Single Cell Sequencing Data

Sequencing results were analyzed using the Cellranger pipeline (10x Genomics) v2.1.0 and custom python scripts written for this project, which are located in the fastqTomat0 GitHub repository (https://github.com/flatironinstitute/fastqToMat0). The reference genome was obtained from Ensembl (Version R64-1-1) as a FASTA file, and the reference annotations were obtained from Ensembl (Version R64-1-1.93) in GTF format. The reference transcript

annotations were altered to incorporate 5' and 3' untranslated regions using data from generated using TIF-seq (Pelechano et al., 2013) and the gffAnnotate.py script from fastqTomat0. The antibiotic resistance marker cassettes was added to both the FASTA and GTF files using command line tools. A STAR reference genome was then created from the modified GTF and FASTA files using cellranger mkref.

Raw sequencing reads were converted into FASTQ files using cellranger mkfastq and a 10x Genomics Index CSV file. These FASTQ reads were then aligned to the reference genome and counted using cellranger count. The FASTQ files for indexes not corresponding to the 10x single-cell transcriptome library were processed with the fastqBCLinker.py script from fastqTomat0, which identifies the genotype for each single-cell read and creates a TSV file mapping cell barcodes to genotypes. The count data from cellranger count and the barcode data from fastqBCLinker.py was combined by the tenXtomatrix.py script from fastqTomat0, which discards doublet single-cell reads and produces a dense TSV matrix of counts per gene per cell that can be imported with python's pandas.read_table() or R's read.table(). This matrix is provided as Supplemental Data 2.

Network Inference Using the Inferelator

Network inference with the Inferelator consists of three major steps; data preprocessing and filtering, estimation of transcription factor activities, and regularized regression. Cross-validation of network inference parameters was performed by randomly selecting half of the genes in the gold standard network and removing them. To prevent circularity, any genes that were used in the gold standard were removed from the prior data during cross-validation; for tests where the gold standard network was also used as a prior, this meant that half of the genes in the gold standard network were retained and defined as the gold standard, and half of the genes in the gold standard network were used as priors. A summary table of the cross-validation results is provided as Supplemental Table 5.

The randomized negative control was performed by randomly reassigning gene names in the prior data. Transcription factor labels and expression values were otherwise unchanged. The simulated negative control was performed using simulated data by constructing a probability distribution for the yeast transcriptome from estimates of absolute mRNA abundances (Lahtvee et al., 2017) and randomly sampling this distribution using the synthesize_data.py script from the fastqToMat0 package. Metadata and total UMI count for each cell were retained in this negative control; only the individual gene counts were synthesized from the simulated control probability distribution.

See below for details on each step of the network inference procedure.

Single-Cell Preprocessing and Filtering

Single cell data was loaded as an integer UMI count matrix (Cells x Genes). Genes with a variance of 0 for all cells were removed. The count matrix was then transformed by log scaling using $\log_2(x+1)$. For data sets that had already undergone library normalization and transformation as a result of an imputation method, this transformation preprocessing step was skipped.

Single-Cell Imputation

All imputation methods used the untransformed integer UMI count matrix (Cells x Genes) in which genes with a variance of 0 had been removed. For MAGIC, count data was library size normalized with the library.size.normalize() function from the Rmagic package, then transformed by square-root, and subjected to imputation with the magic() function from the Rmagic package. For VIPER, count data was subjected directly to the VIPER() function from the VIPER package, using the parameters recommended by the VIPER authors for 10x genomics UMI count data. For ScImpute, count data was normalized by the method included in the ScImpute package and then subjected to imputation with the imputation_wlabel_model8() function from the ScImpute package. The R script to perform these imputations is included with Supplemental Data 1.

Construction of Known Prior TF-Gene Networks

Construction of the gold standard prior network has been previously described (Tchourine et al., 2018); this gold standard network consists of 1,403 signed [-1, 0, 1] interactions, for which sign represents activation (+) or inhibition (-), in a 998 genes by 98 transcription factors regulatory matrix. YEASTRACT priors were retrieved from the YEASTRACT database (Teixeira et al., 2018) using the *generate regulation matrix* tool. Both activation and inhibition interactions were included, but only those that are supported by both DNA binding and expression evidence. The YEASTRACT prior network consists of 9,969 unsigned [0, 1] interactions in a 3,700 genes by 148 transcription factors regulatory matrix. Construction of the ATAC-motif priors has been previously described (Castro et al., 2019; Miraldi et al., 2019), and are built from chromatin accessibility data and known transcription factor binding motifs. The ATAC-motif prior network consists of 71,865 signed integer interactions with a range of [-11, ..., 26], for which sign represents activation (+) or inhibition (-) and absolute values represent the number of motif occurrences, in a 5,551 genes by 138 transcription factors regulatory matrix. Bussemaker-priors were generated from modeling transcription factor affinities for regulatory DNA motifs (Ward and Bussemaker, 2008). The Bussemaker prior network consists of unsigned floating-point values [0, 20] that reflect estimated binding affinities in a dense 6,516 genes by 123 transcription factors regulatory matrix.

Estimating transcription factor activities (TFA)

Log-transformed single-cell data was transposed into matrix $\boldsymbol{X}$, in which columns are individual cells and rows are genes. $\boldsymbol{P}$ is the connectivity matrix of known prior regulatory interactions between transcription factors (in columns) and genes (in rows). $\boldsymbol{P}_{i,k}$ is zero if there is no known

regulatory interaction between transcription factor *k* and gene *i*. ***A*** is the activity matrix, where the columns are the individual cells as in ***X*** and rows are the transcription factors. We model the expression of gene *i* in individual cell *j* as a linear combination of the activities of the *a priori* known regulators of gene *i* in individual cell *j* (1). In practice, this means that we use the known targets of a transcription factor to derive its activity.

$$X_{i,j} = \sum_{k \in TFs} P_{i,k} A_{k,j} \quad (1)$$

In matrix form, equation (1) can be written as $X = PA$. This is an overdetermined system, meaning that there are more equations than unknowns and therefore there is no solution if all equations are linearly independent. We approximate A by finding $\hat{A}$ that minimizes $\| P\hat{A} - X \|_2^2$. If a transcription factor has no prior targets present in ***P***, we use the expression of that transcription factor as a proxy for its activity.

Inferring regulatory interactions, single-task (Bayesian Best Subset Regression)

We utilize a bayesian best-subset regression (BBSR) method, previously described (Greenfield et al., 2013), for single-task network inference. At steady state, we model the expression of a gene *i* in individual cell *j* as a linear combination of the activities of its regulators in individual cell *j* (2). For each gene *i*, we limit the number of potential regulators R_i to the ten with the highest context likelihood of relatedness, calculated from the mutual information between all regulators and the gene *i* (Madar et al., 2010), in addition to any *a priori* known regulator of gene *i*. Limiting the regulators is necessary before best subset regression, when we find the least squares solution to all possible combinations of predictors in set R_i. Because we expect a limited number of transcription factors to regulate a particular gene, our goal is to find a sparse solution for β, in which non-zero entries define both the strength and direction (activation or repression) of a regulatory relationship.

$$X_{i,j} = \sum_{k \in R_i} \beta_{i,k} A_{k,j} \quad (2)$$

Prior knowledge can be incorporated using Zellner's g-prior on the regression parameters β; in this work, we include prior interactions in the set of predictors to be modeled by best subset regression, but we do not further bias the predictors chosen with a g-prior on the regression parameters. We select the model with the lowest Bayesian Information Criterion, which adds a theoretically derived penalty term to the training error to account for model complexity and thereby reduce generalization error. After this step, the output is a matrix of inferred regression parameters β, where each entry corresponds to a regulatory relationship between transcription factor *k* and gene *i*.

Inferring regulatory interactions, multi-task (AMuSR)

The multitask approach used here entails a joint inference of regulatory networks across multiple expression datasets. In addition to the linear assumption, in which gene expression is a linear function of the activities of regulators, we also assume that much of the underlying regulatory network is shared among related datasets (conditions). Here, we extend a previous version of the Inferelator that implements Adaptive Multiple Sparse Regression (AMuSR), which is designed to leverage cross-dataset commonalities while preserving relevant differences (Castro et al., 2019).

There are multiple ways of dividing the existing yeast data into multiple network data subsets, which we refer to as tasks. Within our experimental design, cells are processed and sequenced as batches, which are taken from separate environmental growth conditions. Differences between these batches are a combination of technical and biological variation. The technical variation can come from batch effect due to stress and energy-source differences associated with differing growth conditions (for example via direct effects on cell wall and thus cell lysis/yield), as well as from differences in sample preparation and sequencing. Differences in growth condition also generate biologically significant variation in gene expression due to differences in regulatory program activation. Removing technical variation while retaining biological variation through batch normalization is not feasible, and therefore these individual sample batches from separate growth conditions are taken as individual tasks for the network inference. Thus, the index 'd', below, ranges from 1 to 11 and is an index over the separate datasets corresponding to growth conditions. This separation into tasks results in the joint learning of 11 networks (one for each growth condition), followed by combination into a single global network.

Briefly, the network model is represented as a matrix ***W*** for each target gene (where columns are individual single-cell batches *d* and rows are potential regulators *k*) with signed entries corresponding to strength and type of regulation. We then decompose the model coefficient matrix ***W*** into a dataset-specific component ***S*** and a conserved component ***B*** to enable us to penalize dataset-unique and conserved interactions separately for each target gene; this separation captures differences in regulatory networks across datasets. Specifically, we apply an l_1/l_∞ penalty to the ***B*** component to encourage similarity between network models, and an l_1/l_1 penalty to the other to accommodate differences to ***S*** (Jalali et al., 2010). Regularization parameters λ_s and λ_b, representing the strength of each penalty, were chosen via Extended Bayesian Information Criterion (Chen and Chen, 2008). We can incorporate prior knowledge by using adaptive weights ($\Phi\lambda_s$) when penalizing different coefficients in the l_1/l_1 penalty (Zou, 2006). In this work, however, we chose not to bias predictors to the priors using adaptive weights, and set Φ to 1. For each gene, we minimize the following function (Castro et al., 2019):

$$argmin_{S,B} = \sum_d \| X_i^{(d)} - \hat{A}^{(d)T}(S_{*,d} + B_{*,d})\|_2^2 + \lambda_s \sum_{k,d} |\Phi_{k,d} S_{k,d}| + \lambda_b \|B\|_{1,\infty} \quad (3)$$

$$output: \; W = S + B$$

Ranking interactions and data resampling

Interactions were ranked by both the overall performance of the model for each gene *i* and the proportion of variance explained by each $\beta_{i,k}$. The output of this procedure is a matrix ***S*** where $S_{i,k}$ is the confidence score on the interaction between transcription factor *k* and gene *i*. In order to avoid overfitting and sampling biases, we repeat this procedure *N* times by resampling the input data matrix with replacement. Finally, we rank combine the confidence scores generated by running the above inference procedure on each of the *N* bootstrapped datasets and obtain a final matrix of combined confidence scores for the possible interactions between transcription factors (columns) and genes (rows).

Network Combination

Individual task networks were assembled into a global network by rank combining the confidence scores generated for each possible interaction between transcription factors and genes, obtaining a final matrix of combined confidence scores for the global network. Global interactions were ordered by combined confidence score, and the top interactions were kept to a threshold defined by precision = 0.5, as determined by recovery of the priors.

STATISTICAL ANALYSIS AND DIFFERENTIAL GENE EXPRESSION

To analyze all growth conditions together, the raw single-cell count matrix was normalized using *multiBatchNorm* from the *scater* package in R (McCarthy et al., 2017). In short, this calculates size factors that are used to scale cells from different environmental condition batches so that each batch is of approximately the same mean UMI count. Cells were then library size normalized within batches and the normalized data was log-transformed with $\log_2(x+1)$ to give a transformed and normalized count matrix.

Visualizing Single Cell Expression Data

This normalized count matrix was reduced to 50 principal components by principal component analysis (PCA) with *multiBatchPCA* from the scater package in R. *MultiBatchPCA* is standard PCA with the modification that each environmental condition batch contributes equally to the covariance matrix, even when batches are imbalanced in cell count. These principal components were projected into two dimensional space by Uniform Manifold Approximation and Projection (*UMAP*) and plotted.

To analyze each growth condition separately, the cells corresponding to a growth condition were selected from the raw count matrix, library-size normalized and $\log_2(x+1)$ transformed, and reduced to 50 principal components with PCA. These principal components were projected into two dimensional space by UMAP for plotting, and also used to generate a shared

nearest-neighbor (sNN) graph, which is used to cluster cells using the Louvain clustering method. Each growth condition was processed and plotted separately.

Single-Cell Differential Gene Expression

The transformed, normalized count matrix from the multiBatchNorm function of the scater package in R (described above) was tested for differential gene expression (Fold Change > 1.5, $p < 0.05$, adjusted by the Benjamini & Hochberg method) between the wild-type strain and the transcription factor deletion strains by pairwise Welch's T-Test, with pairwiseTTest from the scran package in R (Lun et al., 2016).

Pseudobulk Differential Gene Expression

The raw, unmodified UMI counts of all cells from each biological replicate (with the same strain barcode) within a specific environmental growth condition were summed, resulting in 72 samples per condition (6 biological replicates for each of the 12 transcription factor deletions). Summed pseudobulk expression data was then tested with DESeq2 (Love et al., 2014) for differential gene expression (Fold Change > 1.5, $p < 0.05$) with no additional processing or normalization.

Gene Ontology

The number of interactions was determined for each gene and each transcription factor. Interactions are considered Learned (new) if they are present in the learned network and not in the prior network; Learned (In Prior) if they are present in the learned network and not in the prior; and, Not Learned (In Prior) if they are present in the prior and not in the learned network. Each gene was mapped to Gene Ontology (GO) slim terms using the YeastGenome slim mapping (https://downloads.yeastgenome.org/curation/literature/go_slim_mapping.tab), which is a curated gene ontology mapping of high-level, broad GO terms. Interactions for all genes annotated with a GO term were summed. The generic terms "biological_process", "not_yet_annotated", and "other" are removed from both target genes and regulatory transcription factors, and the common term "transcription from RNA polymerase II promoter" was removed from regulatory transcription factors GO annotations. The 25 remaining terms with the highest number of learned (new) interactions were plotted separately for both the target genes and the regulatory transcription factors.

DATA AND SOFTWARE AVAILABILITY

Sequencing Data

Raw sequencing data, the output from the cellranger pipeline to count reads, and the output from the fastqToMat0 pipeline to extract and attach genotype metadata to the count matrix are available in NCBI GEO under the accession number GEO: GSE125162.

Single-Cell Processing Pipeline

The cellranger pipeline is available from 10x Genomics under the MIT license (https://github.com/10XGenomics/cellranger). The fastqToMat0 pipeline is available from github (https://github.com/flatironinstitute/fastqToMat0) and is released under the MIT license. Genome sequence and annotations are included as Supplemental Data 4.

Network Inference

The Inferelator is implemented in Python, with dependencies on the widely-distributed scientific packages Numpy, Scipy, Pandas, and Scikit-learn. Scaling to a high-performance computing cluster is implemented with dask. All network inference in this work was performed with Python v2.7.12, Numpy v1.16.2, Pandas v0.24.2, Scikit-learn v0.20.0, Scipy v1.1.0, and dask v1.1.3. The inferelator is available from github (https://github.com/flatironinstitute/inferelator) and is released under the Simplified BSD license.

Figure Construction

Figure 1 and Supplemental Figure 1 are constructed using Adobe Illustrator. Figures 2-7 and Supplemental Figures 2-7 are constructed with R. The R (v3.5.1) packages used are as follows: for plotting, ggplot2 (v3.1.0), cowplot (v0.9.3), ggridges (v0.5.1), viridis (v0.5.1), RColorBrewer (v1.1-2), and scales (v1.0.0); for data manipulation, dplyr (v0.7.7), data.table (v1.11.8), reshape2 (v1.4.3), and stringr (v1.3.1); and for single-cell analysis, scater (v1.10.0), scran (v1.10.1), umap (R: v0.2.0.0, python: v0.3.6), igraph (v1.2.2), and DESeq2 (1.22.2). The R scripts to generate these figures and all required data are included with Supplemental Data 1. Network illustrations in Figure 6 & 7 were generated using Gephi 0.9.2 from the inferelator output network (gefx formatted); the layouts used are Force Atlas 2, Noverlap and Label Adjust. Figures were minimally modified from R outputs to enhance layout and aesthetics using Adobe Illustrator.

Interactive Figures

Interactive versions of several panels from figures 1-4 are available as Shiny (Chang et al., 2018) apps online at http://shiny.bio.nyu.edu/YeastSingleCell2019/. The data used in the interactive figures is provided in Supplemental Data 1. Source code for the Shiny app is available upon request under the MIT license.

Supplemental Data

Supplemental Data 1 is an archive containing R scripts used to generate Figures 2-7 and Supplemental Figures 2-7 with a README detailing the necessary R environment to run them locally. Supplemental Data 1 also contains a **data** folder with the raw count matrix as a TSV file (103118_SS_Data.tsv.gz), the simulated negative data count matrix as a TSV file (110518_SS_NEG_Data.tsv.gz), a gene name metadata TSV file (yeast_gene_names.tsv), supplemental table 5 as a TSV file (STable5.tsv), and the yeast gene ontology slim mapping as a TAB file (go_slim_mapping.tab). Supplemental Data 1 also contains a **priors** folder with the Gold Standard and three sets of priors data tested in this work, all as TSV files. Supplemental Data 1 also contains a **network** folder with the network learned in this paper (signed_network.tsv) as a TSV file, and the networks for each experimental condition (COND_signed_network.tsv) as 11 separate TSV files.

Supplemental Data 2 is the raw count matrix as a TSV file (103118_SS_Data.tsv.gz).

Supplemental Data 3 is the network learned in this paper as a TSV file (network.tsv).

Supplemental Data 4 is an archive containing the sequences used for mapping reads; a FASTA file containing the genotype-specific barcodes (bcdel_1_barcodes.fasta), a FASTA file containing the yeast S288C genome modified with markers (Saccharomyces_cerevisiae.R64-1-1.dna.toplevel.Marker.fa), and a GTF file containing the yeast gene annotations modified to include untranslated regions at the 5' and 3' end, and with markers (Saccharomyces_cerevisiae.R64-1-1.Marker.UTR.notRNA.gtf).

Supplemental Data 5 is an excel file containing Supplemental Tables 1-5.

Supplemental References

Castro DM, de Veaux NR, Miraldi ER, Bonneau R. 2019. Multi-study inference of regulatory networks for more accurate models of gene regulation. *PLoS Comput Biol* **15**:e1006591.

Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. 2018. shiny: Web Application Framework for R. *CRAN*.

Chen J, Chen Z. 2008. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95**:759–771.

Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**:31–34.

Greenfield A, Hafemeister C, Bonneau R. 2013. Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics* **29**:1060–1067.

Jalali A, Sanghavi S, Ruan C, Ravikumar PK. 2010. A Dirty Model for Multi-task Learning In: Lafferty JD, Williams CKI, Shawe-Taylor J, Zemel RS, Culotta A, editors. Advances in Neural Information Processing Systems 23. Curran Associates, Inc. pp. 964–972.

Lahtvee P-J, Sánchez BJ, Smialowska A, Kasvandik S, Elsemman IE, Gatto F, Nielsen J. 2017. Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. *Cell Systems* **4**:495–504.e5.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for

RNA-seq data with DESeq2. *Genome Biol* **15**:550.
Lun ATL, McCarthy DJ, Marioni JC. 2016. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**:2122.
Madar A, Greenfield A, Vanden-Eijnden E, Bonneau R. 2010. DREAM3: Network Inference Using Dynamic Context Likelihood of Relatedness and the Inferelator. *PLoS One* **5**:e9803.
McCarthy DJ, Campbell KR, Lun ATL, Wills QF. 2017. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**:1179–1186.
Miraldi ER, Pokrovskii M, Watters A, Castro DM, De Veaux N, Hall JA, Lee J-Y, Ciofani M, Madar A, Carriero N, Littman DR, Bonneau R. 2019. Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells. *Genome Res*. doi:10.1101/gr.238253.118
Pelechano V, Wei W, Steinmetz LM. 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**:127–131.
Tchourine K, Vogel C, Bonneau R. 2018. Condition-Specific Modeling of Biophysical Parameters Advances Inference of Regulatory Networks. *Cell Rep* **23**:376–388.
Teixeira MC, Monteiro PT, Palma M, Costa C, Godinho CP, Pais P, Cavalheiro M, Antunes M, Lemos A, Pedreira T, Sá-Correia I. 2018. YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in Saccharomyces cerevisiae. *Nucleic Acids Res* **46**:D348–D353.
Ward LD, Bussemaker HJ. 2008. Predicting functional transcription factor binding through alignment-free and affinity-based analysis of orthologous promoter sequences. *Bioinformatics* **24**:i165–71.
Zou H. 2006. The Adaptive Lasso and Its Oracle Properties. *J Am Stat Assoc* **101**:1418–1429.

Supplemental Table 1

Purpose	Primer ID	Name	Sequence	Notes
Construct DGP304	1628	KAN_FWD	ttgtaaaacgacggccagtgcccggggacatggaggcccagaatac	Amplifies pTEF::KAN from pUG6
	1629	KAN_REV	tctagaccagtcttctcgaagacacggatccttagaaaaactcatcgagcatcaaatg	Amplifies pTEF::KAN from pUG6
	1630	tTEF_FWD	ttcgagaagactggtctagaactgacaataaaaagattcttgttttc	Amplifies tTEF from pUG6
	1659	tTEF_REV_2	caagcttgcatgcctgcaggcccgggatcgacagcagtatagcgaccagc	Amplifies tTEF from pUG6
	1641	pUC_LEFT	cactggccgtcgttttacaa	Amplifies pUC19
	1642	pUC_RIGHT	cctgcaggcatgcaagcttg	Amplifies pUC19
Construct DGP306	1751	BC_GFP_F	gctcgatgagtttttctaaggatccgtgtcttcgaaagtgaaacgtgatttc	Amplifies GFP from pWS158
	1752	BC_GFP_R	gaatctttttattgtcagttctagacacgtcttctataaacgcagaaaggcc	Amplifies GFP from pWS158
Construct Barcode	1738	BARCODE_2	GTAAAACGACGGCCAGTgaagacgtatccaNNgNNgtNNgNNgtNNgNNgtctgtgtcttcGTCATAGCTGTTTCCTG	Degenerate Barcode Oligo
	1671	M13R-17	CAGGAAACAGCTATGAC	Double-Stranding Barcode Oligo
Genomic Geletion Primers: FWD & REV Contain Homology Arms, C & D Are Confirmation Primers	1702	RTG1_FWD	AAAACACTAGATAGTGAACCAAAAGAAAGCACAACACCAAACATGcggggacatggaggc	Deletion of RTG1
	1703	RTG1_REV	AGGGTTATCACAACATAGCAATAGTGAGAGTCAGAAGTACTTTTAgggatcgacagcagt	Deletion of RTG1
	1704	RTG1_C	CCAGCGCTGAAGTCGCC	Confirmation of RTG1 Deletion
	1705	RTG1_D	AAACAGTCTTGCATTCATAACATCA	Confirmation of RTG1 Deletion
	1706	RTG3_FWD	TTATTTTTTGTCAGGCGAACCTACTTCTTAAATAAGTGAAGAATGcggggacatggaggc	Deletion of RTG3
	1707	RTG3_REV	CCTTTTTCAAATTTAATTTTTTCCCGCTAATAAGACCATAAACTAgggatcgacagcagt	Deletion of RTG3
	1708	RTG3_C	TATCTCCTTTTACCAATAACCACCA	Confirmation of RTG3 Deletion
	1709	RTG3_D	TCATGAAAGAAAATTACATGGGTTT	Confirmation of RTG3 Deletion
	1710	GLN3_FWD	GAGAGAGAGCACAGGGCCCCCTTTTCCCCCACCAACAAACAAATGcggggacatggaggc	Deletion of GLN3
	1711	GLN3_REV	AATTATTAACATAATAAGAATAATGATAATGATAATACGCGGTCAgggatcgacagcagt	Deletion of GLN3
	1712	GLN3_C	TTATATCGTCCGCAAACTTTAACTC	Confirmation of GLN3 Deletion
	1713	GNL3_D	AAACAAATAATACCAATGCTCAGGA	Confirmation of GLN3 Deletion
	1714	GAT1_FWD	ACATATATATAGGTGTGTGCCACTCCCGGCCCCGGTATTAGCATGcggggacatggaggc	Deletion of GAT1
	1715	GAT1_REV	GGGGGATCTATAAATTCAGATTCAACCAATCCAGGCTCAGATTGCgggatcgacagcagt	Deletion of GAT1
	1716	GAT1_C	TAGATCATCATTCTGGAGACACAAA	Confirmation of GAT1 Deletion
	1717	GAT1_D	GGTTGCTCTTACTGAGCTTCTAGTG	Confirmation of GAT1 Deletion
	1718	DAL80_FWD	CTATATTACATTATATTATATCATATTATATATAAGAGAATGGTGcggggacatggaggc	Deletion of DAL80
	1719	DAL80_REV	CTATATCCTGTTATATTTATACTGTCATGAACGCTACATCCTTTCgggatcgacagcagt	Deletion of DAL80
	1720	DAL80_C	GAAAGAAGAGACTGCTTACAACTGG	Confirmation of DAL80 Deletion
	1721	DAL80_D	CTTCATTTTCATTAAGAGCACGTTT	Confirmation of DAL80 Deletion
	1722	DAL81_FWD	TGTTTAGACGAGCGGCAGAACGACAGGCAGCCATACTATCAAATGcggggacatggaggc	Deletion of DAL81
	1723	DAL81_REV	TTATTACCGTATTCCATTTTTACTTATGTGCTATTATTTATATTAgggatcgacagcagt	Deletion of DAL81
	1724	DAL81_C	GAAGTCCAGCCATAAATAGTGCTAA	Confirmation of DAL81 Deletion
	1725	DAL81_D	GCAAAGTATAATAGACGAGGCAAAA	Confirmation of DAL81 Deletion
	1726	DAL82_FWD	GGTTTAACGATACTTTGAAAGGTTTAGGCGGAGCCAACCATAATGcggggacatggaggc	Deletion of DAL82
	1727	DAL82_REV	AGATACATATGGACATGAATCAAGCATATCTAATAGAAGTTTTTAgggatcgacagcagt	Deletion of DAL82
	1728	DAL82_C	ACCATATGCCTACAGACAAGATAGC	Confirmation of DAL82 Deletion
	1729	DAL82_D	CTTGGTGAAATGGAAGGTAAATATG	Confirmation of DAL82 Deletion
	1730	GZF3_FWD	TCGGGTATTTGTACTGTGCTCTACAATTCTACTGGTAATATTATGcggggacatggaggc	Deletion of GZF3
	1731	GZF3_REV	ATTATATTTGAAAAATGATGCTGATTTTTTTGAAAAAAAGATCTAgggatcgacagcagt	Deletion of GZF3
	1732	GZF3_C	ACATATCAATGAATTAGACGGGAAA	Confirmation of GZF3 Deletion
	1733	GZF3_D	CTTTTTGCAGTAGTTATTTGCATCA	Confirmation of GZF3 Deletion
	1734	GCN4_FWD	CACCAGCCACACAGCTCACTCATCTACTTCGCAATCAAAACAAAAcggggacatggaggc	Deletion of GCN4+ Upstream uORFs
	1735	GCN4_REV	ACGAGAATGAAATAAAAAATATAAAATAAAAGGTAAATGAAATCAgggatcgacagcagt	Deletion of GCN4+ Upstream uORFs
	1736	GCN4_C	CAAACAAGAAAGGTTAAGAAACCAA	Confirmation of GCN4 Deletion
	1737	GCN4_D	TTTAAAGTTTCATTCCAGCATTAGC	Confirmation of GCN4 Deletion
	1804	STP1_FWD	TATATGTACATAGACATATAAAGGGCATGTGGTAATTATGACATGcggggacatggaggc	Deletion of STP1
	1805	STP1_REV	TTCCAATATGATACCCTTATTTTTATCCCGTGTTATATTTAATCAgggatcgacagcagt	Deletion of STP1
	1806	STP1_C	ATCTTCATATGATTTGAGCGAAGAC	Confirmation of STP1 Deletion
	1807	STP1_D	AAAAAGAAATCACAAACGTCAAAAG	Confirmation of STP1 Deletion
	1808	STP2_FWD	AAAATACCTGAAACCGCCATAAAAATAATACCTGAATTAACCTTAcggggacatggaggc	Deletion of STP2
	1809	STP2_REV	GCAAGAGAGCAAAAAGTTGGGGAAGTAAGGAAATCCATTATCATGgggatcgacagcagt	Deletion of STP2
	1853	STP2_C	GCTCCTCTAAGGGCTAGAATAGAGA	Confirmation of STP2 Deletion
	1854	STP2_D	TTTGTTGTTCTGTTTGGGAAATAAT	Confirmation of STP2 Deletion
	1918	HO_FWD	CATATCCTCATAAGCAGCAATCAATTCTATCTATACTTTAAAATGcggggacatggaggc	Deletion of HO (WT Control)
	1919	HO_REV	TTACTTTTATTACATACAACTTTTTAAACTAATATACACATTTTAgggatcgacagcagt	Deletion of HO (WT Control)
	1888	HO_C	GAGTGGTAAAAATCGAGTATGTGCT	Confirmation of HO Deletion
	1889	HO_D	GTCACTGACCCAGTCTTGTC	Confirmation of HO Deletion
Colony PCR	47	KanC	TGATTTTGATGACGAGCGTAAT	Confirmation PCR Primer for KanMX
	613	NatC	TTTTCGCCTCGACATCATCT	Confirmation PCR Primer for NatMX
Seq.	1634	KAN_3SEQ_FWD	CAGACCGATACCAGGATCTTG	Sanger Sequencing Primer for Barcode Recovery
Genotype Barcode PCR and Indexing	1949	10x_Dialout_F	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNgctcgatgagtttttctaaggatcc	Amplifies out KanBC barcodes from 10x WTA.
	1950	10x_Dialout_R	CTACACTCTTTCCCTACACGACGCTCTTCCG	Amplifies out KanBC barcodes from 10x WTA.
	1951	10x_PD_D701_P7	CAAGCAGAAGACGGCATACGAGATcgagtaatGTGACTGGAGTTCAGACGTGT	Adds P7 & D701 i7 sequence to PERTURB dialout
	1952	10x_PD_D702_P7	CAAGCAGAAGACGGCATACGAGATtctccggaGTGACTGGAGTTCAGACGTGT	Adds P7 & D702 i7 sequence to PERTURB dialout
	1953	10x_PD_D704_p7	CAAGCAGAAGACGGCATACGAGATgagattccGTGACTGGAGTTCAGACGTGT	Adds P7 & D704 i7 sequence to PERTURB dialout
	1954	10x_PD_P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG	Adds P5 sequence to PERTURB dialout

Supplemental Table 2

Strain ID	Strain	Genotype	Parent	Barcode	Plate Location	Reference
1	FY4	MATa				
2	FY5	MATα				
1997	WT(ho)_A1	MATa Δho::KanMX	FY4	CAGAGGTTCGTCGTACGGC		This work
2062	WT(ho)_B1	MATa Δho::KanMX	FY4	TAGGTGTTTGATGTAAGTT		This work
2063	WT(ho)_C1	MATa Δho::KanMX	FY4	TAGAAGTTGGCAGTGCGTA		This work
2112	WT(ho)_D1	MATa Δho::KanMX	FY4	TCGGTGTGTGTTGTGGGTC		This work
2113	WT(ho)_E1	MATa Δho::KanMX	FY4	CTGTAGTGCGGAGTCCGAG		This work
2114	WT(ho)_F1	MATa Δho::KanMX	FY4	TAGATGTGTGGGGTAAGCC		This work
1840	rtg1_B2	MATa Δrtg1::KanMX	FY4	CGGTCGTTGGTAGTTAGTG		This work
1899	rtg1_C2	MATa Δrtg1::KanMX	FY4	CAGGTGTATGTAGTGGGTA		This work
1900	rtg1_D2	MATa Δrtg1::KanMX	FY4	TCGATGTCCGAAGTAGGAC		This work
1901	rtg1_E2	MATa Δrtg1::KanMX	FY4	AGGGGGTTGGCGGTGGGGC		This work
1902	rtg1_F2	MATa Δrtg1::KanMX	FY4	GAGGCGTGGGGGGTTCGGC		This work
1984	rtg1_G2	MATa Δrtg1::KanMX	FY4	GAGTAGTAAGCTGTAAGCG		This work
1841	rtg3_C3	MATa Δrtg3::KanMX	FY4	CTGTAGTCGGGCGTCTGTG		This work
1933	rtg3_D3	MATa Δrtg3::KanMX	FY4	TGGTTGTGTGTAGTACGTT		This work
1934	rtg3_E3	MATa Δrtg3::KanMX	FY4	TCGCAGTGAGATGTGTGTC		This work
1935	rtg3_F3	MATa Δrtg3::KanMX	FY4	CGGGGGTGGGCCGTGGGTG		This work
1936	rtg3_G3	MATa Δrtg3::KanMX	FY4	CCGTGGTAAGTAGTAAGTG		This work
1937	rtg3_H3	MATa Δrtg3::KanMX	FY4	TCGTAGTTAGTTGTAAGAT		This work
1839	gln3_A4	MATa Δgln3::KanMX	FY4	AAGCAGTATGTAGTGAGTC		This work
1893	gln3_D4	MATa Δgln3::KanMX	FY4	GTGCGGTTGGGCGTGGGCG		This work
1894	gln3_E4	MATa Δgln3::KanMX	FY4	CGGAAGTTTGCCGTTGGCC		This work
1895	gln3_F4	MATa Δgln3::KanMX	FY4	GTGGCGTGTGCGGTGCGGG		This work
1896	gln3_G4	MATa Δgln3::KanMX	FY4	CAGACGTGGGTAGTGAGAA		This work
1897	gln3_H4	MATa Δgln3::KanMX	FY4	TTGACGTCAGCCGTACGGA		This work
1838	gat1_A5	MATa Δgat1::KanMX	FY4	GCGGAGTCCGTGGTTAGGG		This work
1846	gat1_B5	MATa Δgat1::KanMX	FY4	AAGCCGTTTGGGGTGCGCT		This work
1847	gat1_E5	MATa Δgat1::KanMX	FY4	TCGACGTTGGAGGTTGGGG		This work
1848	gat1_F5	MATa Δgat1::KanMX	FY4	TTGTTGTGGGTCGTCCGAA		This work
1849	gat1_G5	MATa Δgat1::KanMX	FY4	AAGTTGTCGGCAGTGGGTT		This work
1850	gat1_H5	MATa Δgat1::KanMX	FY4	TGGGAGTGCGCCGTCGGCG		This work
1837	dal80_A6	MATa Δdal80::KanMX	FY4	GCGTTGTAAGGTGTCGGTC		This work
1865	dal80_B6	MATa Δdal80::KanMX	FY4	CTGCGGTTCGGTGTGAGTG		This work
1866	dal80_C6	MATa Δdal80::KanMX	FY4	AAGGTGTACGAAGTGTGAT		This work
1867	dal80_F6	MATa Δdal80::KanMX	FY4	AGGAAGTCTGCTGTGCGTA		This work
1868	dal80_G6	MATa Δdal80::KanMX	FY4	CGGCAGTTAGTAGTACGAT		This work
1982	dal80_H6	MATa Δdal80::KanMX	FY4	ACGCAGTACGATGTTTGCA		This work
1842	dal81_A7	MATa Δdal81::KanMX	FY4	GGGGCGTGAGGGGTGGGCT		This work
1869	dal81_B7	MATa Δdal81::KanMX	FY4	TAGTAGTGTGTAGTCTGTA		This work
1870	dal81_C7	MATa Δdal81::KanMX	FY4	CGGCAGTATGACGTTAGTA		This work
1871	dal81_D7	MATa Δdal81::KanMX	FY4	TTGGCGTCGGATGTAGGTC		This work
1872	dal81_G7	MATa Δdal81::KanMX	FY4	GGGAAGTCTGTAGTGTGCT		This work
1873	dal81_H7	MATa Δdal81::KanMX	FY4	TAGAGGTTCGATGTATGAT		This work
1843	dal82_A8	MATa Δdal82::KanMX	FY4	ATGCAGTACGGTGTCGGCA		This work
1876	dal82_B8	MATa Δdal82::KanMX	FY4	TGGAGGTTTGAAGTCAGAT		This work
1877	dal82_C8	MATa Δdal82::KanMX	FY4	GAGATGTTAGAGGTTCGGT		This work
1878	dal82_D8	MATa Δdal82::KanMX	FY4	AAGAAGTTGGTTGTCAGGG		This work
1879	dal82_E8	MATa Δdal82::KanMX	FY4	TCGCGGTCAGTAGTTTGAC		This work
1880	dal82_H8	MATa Δdal82::KanMX	FY4	CAGAAGTGCGTTGTTTGAT		This work
1844	gzf3_A9	MATa Δgzf3::KanMX	FY4	GTGCGGTTTGACGTGTGCC		This work
1858	gzf3_B9	MATa Δgzf3::KanMX	FY4	CAGTCGTTAGATGTAGGGG		This work
1859	gzf3_C9	MATa Δgzf3::KanMX	FY4	ACGGAGTAAGTGGTCAGTT		This work
1860	gzf3_D9	MATa Δgzf3::KanMX	FY4	ACGCTGTTTGATGTTAGTT		This work
1861	gzf3_E9	MATa Δgzf3::KanMX	FY4	CAGTTGTGTGATGTCGGGG		This work
1862	gzf3_F9	MATa Δgzf3::KanMX	FY4	TTGAGGTCTGCAGTGTGTA		This work
1845	gcn4_A10	MATa Δgcn4::KanMX	FY4	AGGTCGTTTGTAGTAGGAC		This work
1851	gcn4_C10	MATa Δgcn4::KanMX	FY4	TGGTGGTAAGAGGTCAGTG		This work
1852	gcn4_D10	MATa Δgcn4::KanMX	FY4	TTGACGTCAGCCGTCCGTG		This work
1853	gcn4_E10	MATa Δgcn4::KanMX	FY4	GAGATGTGTGTGGTTTGAA		This work
1854	gcn4_F10	MATa Δgcn4::KanMX	FY4	ATGGCGTCAGGCGTTCGAT		This work
1855	gcn4_G10	MATa Δgcn4::KanMX	FY4	TAGTAGTAAGGCGTAAGGG		This work
1903	stp1_C11	MATa Δstp1::KanMX	FY4	GTGGTGTTTGACGTAAGGC		This work
1904	stp1_D11	MATa Δstp1::KanMX	FY4	TAGTTGTTGGAGGTATGAT		This work
1905	stp1_E11	MATa Δstp1::KanMX	FY4	TTGGTGTCAGCTGTGGGAG		This work
1925	stp1_F11	MATa Δstp1::KanMX	FY4	GGGTGGTTAGAAGTACGCC		This work
1926	stp1_G11	MATa Δstp1::KanMX	FY4	GTGCGGTGGGCCGTGGGGG		This work
1927	stp1_H11	MATa Δstp1::KanMX	FY4	CAGCCGTTGGTGGTTTGAT		This work
1912	stp2_A12	MATa Δstp2::KanMX	FY4	TTGTAGTCTGTGGTAAGAA		This work
1913	stp2_D12	MATa Δstp2::KanMX	FY4	CTGTTGTGTGAGGTTCGGT		This work
1914	stp2_E12	MATa Δstp2::KanMX	FY4	TTGAAGTTAGAAGTAGGCA		This work
1915	stp2_F12	MATa Δstp2::KanMX	FY4	ATGGAGTTTGTTGTACGCT		This work
1916	stp2_G12	MATa Δstp2::KanMX	FY4	TAGATGTTGGGAGTAGGCA		This work
1978	stp2_H12	MATa Δstp2::KanMX	FY4	CCGAGGTATGTCGTTCGGT		This work
2100	rtg1	MATα Δrtg1::NatMX	FY5			This work
2101	rtg3	MATα Δrtg3::NatMX	FY5			This work
2102	gln3	MATα Δgln3::NatMX	FY5			This work
2103	gat1	MATα Δgat1::NatMX	FY5			This work
2104	dal80	MATα Δdal80::NatMX	FY5			This work
2105	dal81	MATα Δdal81::NatMX	FY5			This work
2106	dal82	MATα Δdal82::NatMX	FY5			This work
2107	gzf3	MATα Δgzf3::NatMX	FY5			This work
2108	gcn4	MATα Δgcn4::NatMX	FY5			This work
2109	stp1	MATα Δstp1::NatMX	FY5			This work
2110	stp2	MATα Δstp2::NatMX	FY5			This work
2111	ho	MATα Δho::NatMX	FY5			This work
N/A	WT(ho)_A1	MATa/MATα Δho::KanMX/Δho::NatMX		CAGAGGTTCGTCGTACGGC	BC_DEL_1_A1	This work

N/A	WT(ho)_B1	MATa/MATα Δho::KanMX/Δho::NatMX	TAGGTGTTTGATGTAAGTT	BC_DEL_1_B1	This work
N/A	WT(ho)_C1	MATa/MATα Δho::KanMX/Δho::NatMX	TAGAAGTTGGCAGTGCGTA	BC_DEL_1_C1	This work
N/A	WT(ho)_D1	MATa/MATα Δho::KanMX/Δho::NatMX	TCGGTGTGTGTTGTGGGTC	BC_DEL_1_D1	This work
N/A	WT(ho)_E1	MATa/MATα Δho::KanMX/Δho::NatMX	CTGTAGTGCGGAGTCCGAG	BC_DEL_1_E1	This work
N/A	WT(ho)_F1	MATa/MATα Δho::KanMX/Δho::NatMX	TAGATGTGTGGGGTAAGCC	BC_DEL_1_F1	This work
N/A	rtg1_B2	MATa/MATα Δrtg1::KanMX/Δrtg1::NatMX	CGGTCGTTGGTAGTTAGTG	BC_DEL_1_B2	This work
N/A	rtg1_C2	MATa/MATα Δrtg1::KanMX/Δrtg1::NatMX	CAGGTGTATGTAGTGGGTA	BC_DEL_1_C2	This work
N/A	rtg1_D2	MATa/MATα Δrtg1::KanMX/Δrtg1::NatMX	TCGATGTCCGAAGTAGGAC	BC_DEL_1_D2	This work
N/A	rtg1_E2	MATa/MATα Δrtg1::KanMX/Δrtg1::NatMX	AGGGGGTTGGCGGTGGGGC	BC_DEL_1_E2	This work
N/A	rtg1_F2	MATa/MATα Δrtg1::KanMX/Δrtg1::NatMX	GAGGCGTGGGGGGTTCGGC	BC_DEL_1_F2	This work
N/A	rtg1_G2	MATa/MATα Δrtg1::KanMX/Δrtg1::NatMX	GAGTAGTAAGCTGTAAGCG	BC_DEL_1_G2	This work
N/A	rtg3_C3	MATa/MATα Δrtg3::KanMX/Δrtg3::NatMX	CTGTAGTCGGGCGTCTGTG	BC_DEL_1_C3	This work
N/A	rtg3_D3	MATa/MATα Δrtg3::KanMX/Δrtg3::NatMX	TGGTTGTGTGTAGTACGTT	BC_DEL_1_D3	This work
N/A	rtg3_E3	MATa/MATα Δrtg3::KanMX/Δrtg3::NatMX	TCGCAGTGAGATGTGTGTC	BC_DEL_1_E3	This work
N/A	rtg3_F3	MATa/MATα Δrtg3::KanMX/Δrtg3::NatMX	CGGGGGTGGGCCGTGGGTG	BC_DEL_1_F3	This work
N/A	rtg3_G3	MATa/MATα Δrtg3::KanMX/Δrtg3::NatMX	CCGTGGTAAGTAGTAAGTG	BC_DEL_1_G3	This work
N/A	rtg3_H3	MATa/MATα Δrtg3::KanMX/Δrtg3::NatMX	TCGTAGTTAGTTGTAAGAT	BC_DEL_1_H3	This work
N/A	gln3_A4	MATa/MATα Δgln3::KanMX/Δgln3::NatMX	AAGCAGTATGTAGTGAGTC	BC_DEL_1_A4	This work
N/A	gln3_D4	MATa/MATα Δgln3::KanMX/Δgln3::NatMX	GTGCGGTTGGGCGTGGGCG	BC_DEL_1_D4	This work
N/A	gln3_E4	MATa/MATα Δgln3::KanMX/Δgln3::NatMX	CGGAAGTTTGCCGTTGGCC	BC_DEL_1_E4	This work
N/A	gln3_F4	MATa/MATα Δgln3::KanMX/Δgln3::NatMX	GTGGCGTGTGCGGTGCGGG	BC_DEL_1_F4	This work
N/A	gln3_G4	MATa/MATα Δgln3::KanMX/Δgln3::NatMX	CAGACGTGGGTAGTGAGAA	BC_DEL_1_G4	This work
N/A	gln3_H4	MATa/MATα Δgln3::KanMX/Δgln3::NatMX	TTGACGTCAGCCGTACGGA	BC_DEL_1_H4	This work
N/A	gat1_A5	MATa/MATα Δgat1::KanMX/Δgat1::NatMX	GCGGAGTCCGTGGTTAGGG	BC_DEL_1_A5	This work
N/A	gat1_B5	MATa/MATα Δgat1::KanMX/Δgat1::NatMX	AAGCCGTTTGGGGTGCGCT	BC_DEL_1_B5	This work
N/A	gat1_E5	MATa/MATα Δgat1::KanMX/Δgat1::NatMX	TCGACGTTGGAGGTTGGGG	BC_DEL_1_E5	This work
N/A	gat1_F5	MATa/MATα Δgat1::KanMX/Δgat1::NatMX	TTGTTGTGGGTCGTCCGAA	BC_DEL_1_F5	This work
N/A	gat1_G5	MATa/MATα Δgat1::KanMX/Δgat1::NatMX	AAGTTGTCGGCAGTGGGTT	BC_DEL_1_G5	This work
N/A	gat1_H5	MATa/MATα Δgat1::KanMX/Δgat1::NatMX	TGGGAGTGCGCCGTCGGCG	BC_DEL_1_H5	This work
N/A	dal80_A6	MATa/MATα Δdal80::KanMX/Δdal80::NatMX	GCGTTGTAAGGTGTCGGTC	BC_DEL_1_A6	This work
N/A	dal80_B6	MATa/MATα Δdal80::KanMX/Δdal80::NatMX	CTGCGGTTCGGTGTGAGTG	BC_DEL_1_B6	This work
N/A	dal80_C6	MATa/MATα Δdal80::KanMX/Δdal80::NatMX	AAGGTGTACGAAGTGTGAT	BC_DEL_1_C6	This work
N/A	dal80_F6	MATa/MATα Δdal80::KanMX/Δdal80::NatMX	AGGAAGTCTGCTGTGCGTA	BC_DEL_1_F6	This work
N/A	dal80_G6	MATa/MATα Δdal80::KanMX/Δdal80::NatMX	CGGCAGTTAGTAGTACGAT	BC_DEL_1_G6	This work
N/A	dal80_H6	MATa/MATα Δdal80::KanMX/Δdal80::NatMX	ACGCAGTACGATGTTTGCA	BC_DEL_1_H6	This work
N/A	dal81_A7	MATa/MATα Δdal81::KanMX/Δdal81::NatMX	GGGGCGTGAGGGGTGGGCT	BC_DEL_1_A7	This work
N/A	dal81_B7	MATa/MATα Δdal81::KanMX/Δdal81::NatMX	TAGTAGTGTGTAGTCTGTA	BC_DEL_1_B7	This work
N/A	dal81_C7	MATa/MATα Δdal81::KanMX/Δdal81::NatMX	CGGCAGTATGACGTTAGTA	BC_DEL_1_C7	This work
N/A	dal81_D7	MATa/MATα Δdal81::KanMX/Δdal81::NatMX	TTGGCGTCGGATGTAGGTC	BC_DEL_1_D7	This work
N/A	dal81_G7	MATa/MATα Δdal81::KanMX/Δdal81::NatMX	GGGAAGTCTGTAGTGTGCT	BC_DEL_1_G7	This work
N/A	dal81_H7	MATa/MATα Δdal81::KanMX/Δdal81::NatMX	TAGAGGTTCGATGTATGAT	BC_DEL_1_H7	This work
N/A	dal82_A8	MATa/MATα Δdal82::KanMX/Δdal82::NatMX	ATGCAGTACGGTGTCGGCA	BC_DEL_1_A8	This work
N/A	dal82_B8	MATa/MATα Δdal82::KanMX/Δdal82::NatMX	TGGAGGTTTGAAGTCAGAT	BC_DEL_1_B8	This work
N/A	dal82_C8	MATa/MATα Δdal82::KanMX/Δdal82::NatMX	GAGATGTTAGAGGTTCGGT	BC_DEL_1_C8	This work
N/A	dal82_D8	MATa/MATα Δdal82::KanMX/Δdal82::NatMX	AAGAAGTTGGTTGTCAGGG	BC_DEL_1_D8	This work
N/A	dal82_E8	MATa/MATα Δdal82::KanMX/Δdal82::NatMX	TCGCGGTCAGTAGTTTGAC	BC_DEL_1_E8	This work
N/A	dal82_H8	MATa/MATα Δdal82::KanMX/Δdal82::NatMX	CAGAAGTGCGTTGTTTGAT	BC_DEL_1_H8	This work
N/A	gzf3_A9	MATa/MATα Δgzf3::KanMX/Δgzf3::NatMX	GTGCGGTTTGACGTGTGCC	BC_DEL_1_A9	This work
N/A	gzf3_B9	MATa/MATα Δgzf3::KanMX/Δgzf3::NatMX	CAGTCGTTAGATGTAGGGG	BC_DEL_1_B9	This work
N/A	gzf3_C9	MATa/MATα Δgzf3::KanMX/Δgzf3::NatMX	ACGGAGTAAGTGGTCAGTT	BC_DEL_1_C9	This work
N/A	gzf3_D9	MATa/MATα Δgzf3::KanMX/Δgzf3::NatMX	ACGCTGTTTGATGTTAGTT	BC_DEL_1_D9	This work
N/A	gzf3_E9	MATa/MATα Δgzf3::KanMX/Δgzf3::NatMX	CAGTTGTGTGATGTCGGGG	BC_DEL_1_E9	This work
N/A	gzf3_F9	MATa/MATα Δgzf3::KanMX/Δgzf3::NatMX	TTGAGGTCTGCAGTGTGTA	BC_DEL_1_F9	This work
N/A	gcn4_A10	MATa/MATα Δgcn4::KanMX/Δgcn4::NatMX	AGGTCGTTTGTAGTAGGAC	BC_DEL_1_A10	This work
N/A	gcn4_C10	MATa/MATα Δgcn4::KanMX/Δgcn4::NatMX	TGGTGGTAAGAGGTCAGTG	BC_DEL_1_C10	This work
N/A	gcn4_D10	MATa/MATα Δgcn4::KanMX/Δgcn4::NatMX	TTGACGTCAGCCGTCCGTG	BC_DEL_1_D10	This work
N/A	gcn4_E10	MATa/MATα Δgcn4::KanMX/Δgcn4::NatMX	GAGATGTGTGTGGTTTGAA	BC_DEL_1_E10	This work
N/A	gcn4_F10	MATa/MATα Δgcn4::KanMX/Δgcn4::NatMX	ATGGCGTCAGGCGTTCGAT	BC_DEL_1_F10	This work
N/A	gcn4_G10	MATa/MATα Δgcn4::KanMX/Δgcn4::NatMX	TAGTAGTAAGGCGTAAGGG	BC_DEL_1_G10	This work
N/A	stp1_C11	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	GTGGTGTTTGACGTAAGGC	BC_DEL_1_C11	This work
N/A	stp1_D11	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	TAGTTGTTGGAGGTATGAT	BC_DEL_1_D11	This work
N/A	stp1_E11	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	TTGGTGTCAGCTGTGGGAG	BC_DEL_1_E11	This work
N/A	stp1_F11	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	GGGTGGTTAGAAGTACGCC	BC_DEL_1_F11	This work
N/A	stp1_G11	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	GTGCGGTGGGCCGTGGGGG	BC_DEL_1_G11	This work
N/A	stp1_H11	MATa/MATα Δstp2::KanMX/Δstp2::NatMX	CAGCCGTTGGTGGTTTGAT	BC_DEL_1_H11	This work
N/A	stp2_A12	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	TTGTAGTCTGTGGTAAGAA	BC_DEL_1_A12	This work
N/A	stp2_D12	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	CTGTTGTGTGAGGTTCGGT	BC_DEL_1_D12	This work
N/A	stp2_E12	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	TTGAAGTTAGAAGTAGGCA	BC_DEL_1_E12	This work
N/A	stp2_F12	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	ATGGAGTTTGTTGTACGCT	BC_DEL_1_F12	This work
N/A	stp2_G12	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	TAGATGTTGGGAGTAGGCA	BC_DEL_1_G12	This work
N/A	stp2_H12	MATa/MATα Δstp1::KanMX/Δstp1::NatMX	CCGAGGTATGTCGTTCGGT	BC_DEL_1_H12	This work

Supplemental Table 3

Plasmid ID (DGP)	Name	Description	Source
52	pUC19		
212	pUG6	loxP-flanked marker gene deletion cassette: loxP-pAgTEF1-kanMX-tAgTEF1-loxP, selectable phenotype: G418 resistance	Euroscarf
290	pWS158		Addgene
304	pUC19-KanBC	DG1628+1629 from pUG6 (pTEF-KAN) and DG1630+1659 from pUG6 (tTEF) assembled into pUC19 backbone (amplified with DG1640+1641)	This work
306	pUC19-KanBC[GFP]	DGP304 linearized with BamHI & XbaI and GFP amplified from DGP290 with DG1751 & 1752 assembled by gibson	This work

Supplemental Table 4

Media	Composition	Source
YPD	2% (w/v) peptone	BD Difco #DF0118-17-0
	2% (w/v) glucose	BD Difco #DF0155-08-5
	1% (w/v) yeast extract	BD Difco #DF0127-17-9
YPEthanol	2% (w/v) peptone	BD Difco #DF0118-17-0
	2% (v/v) ethanol	Pharmco-Aaper #111000200
	1% (w/v) yeast extract	BD Difco #DF0127-17-9
Minimal Trace Metals (1x)	500 µg/L (8.09 µM) Boric Acid	Sigma #B6768
	400 µg/L (2.37 µM) Manganese (II) Sulfate Monohydrate	Sigma # M7634
	400 µg/L (1.39 µM) Zinc Sulfate Heptahydrate	Fisher #Z68-500
	200 µg/L (0.83 µM) Sodium Molybdate Dihydrate	Sigma #331058
	200 µg/L (0.74 µM) Ferric (III) Chloride Hexahydrate	Fisher #I88-100
	100 µg/L (0.60 µM) Potassium Iodide	Sigma #60400
	40 µg/L (0.16 µM) Copper (II) Sulfate Pentahydrate	Sigma #209198
Minimal Vitamins (1x)	2 mg/L (11.10 µM) Inositol	Acros Organics #AC12226-1000
	400 µg/L (3.25 µM) Nicotinic Acid	Sigma #N4126
	400 µg/L (1.95 µM) Pyridoxine Hydrochloride	Sigma #P9755
	200 µg/L (1.46 µM) 4-Aminobenzoic Acid	Acros Organics #AC14621-2500
	400 µg/L (1.19 µM) Thiamine Hydrochloride	Fisher #BP892-100
	400 µg/L (0.84 µM) D-Calcium Pantothenate	Acros Organics #AC24330-1000
	200 µg/L (0.53 µM) Riboflavin	Sigma #R4500-25G
	2 µg/L (8.19 nM) D-Biotin	Fisher #BP232-1
	2 µg/L (4.53 nM) Folic Acid	Sigma #F7876
Minimal Salts (1x)	0.1 g/L (0.68 mM) Calcium Chloride Dihydrate	Fisher #C79-500
	0.1 g/L (1.71 mM) Sodium Chloride	Fisher #BP358-212
	0.5 g/L (2.46 mM) Magnesium Chloride Hexahydrate	Sigma #M9272
	1.0 g/L (7.35 mM) Potassium Phosphate Monobasic	VWR #97062-346
Minimal Media	1x Minimal Trace Metals	Above
	1x Minimal Salts	Above
	1x Minimal Vitamins	Above
	2% (w/v) glucose	BD Difco #DF0155-08-5
	20 mM Ammonium Sulfate	Acros Organics #AC423400010
Minimal Media (Ethanol)	1x Minimal Trace Metals	Above
	1x Minimal Salts	Above
	1x Minimal Vitamins	Above
	2% (v/v) ethanol	Pharmco-Aaper #111000200
	20 mM Ammonium Sulfate	Acros Organics #AC423400010
Nitrogen Limiting Minimal Media (Glutamine)	1x Minimal Trace Metals	Above
	1x Minimal Salts	Above
	1x Minimal Vitamins	Above
	2% (w/v) glucose	BD Difco #DF0155-08-5
	800 µM L(+)-Glutamine	Acros Organics #119951000
Nitrogen Limiting Minimal Media (Proline)	1x Minimal Trace Metals	Above
	1x Minimal Salts	Above
	1x Minimal Vitamins	Above
	2% (w/v) glucose	BD Difco #DF0155-08-5
	800 µM L-Proline	Sigma #P0380-100G
Nitrogen Limiting Minimal Media (Ammonium Sulfate)	1x Minimal Trace Metals	Above
	1x Minimal Salts	Above
	1x Minimal Vitamins	Above
	2% (w/v) glucose	BD Difco #DF0155-08-5
	800 µM Ammonium Sulfate	Acros Organics #AC423400010
Nitrogen Limiting Minimal Media (Ammonium Sulfate)	1x Minimal Trace Metals	Above
	1x Minimal Salts	Above
	1x Minimal Vitamins	Above
	2% (w/v) glucose	BD Difco #DF0155-08-5
	800 µM Urea	Acros Organics #AC434720010
Starvation Media	1x Minimal Trace Metals	Above
	1x Minimal Salts	Above
	1x Minimal Vitamins	Above
	1 mM Ammonium Sulfate	Acros Organics #AC423400010