An inducible CRISPR-interference library for genetic interrogation of *Saccharomyces cerevisiae* biology

Amir Momen-Roknabadi[1,2,3], Panos Oikonomou[1,2,3], Maxwell Zegans[2], Saeed Tavazoie[1,2,3*]

[1]Department of Biological Sciences, Columbia University, New York City, New York, United States

[2]Department of Systems Biology, Columbia University, New York City, New York, United States

[3]Department of Biochemistry and Molecular Biophysics, Columbia University, New York City, New York, United States

* E-mail: st2744@columbia.edu

Abstract

Genome-scale CRISPR interference (CRISPRi) is widely utilized to study cellular processes in a variety of organisms. Despite its dominance as a model eukaryote, a genome-wide CRISPRi library, optimized for targeting the *Saccharomyces cerevisiae* genome, has not been presented to date. We have generated a comprehensive, inducible CRISPRi library, based on spacer design rules optimized for yeast. We have validated this library for genome-wide interrogation of gene function across a variety of applications, including accurate discovery of haploinsufficient genes and identification of enzymatic and regulatory genes involved in adenine and arginine biosynthesis. CRISPRi screens, using this library can identify genes and pathways with high precision and low false discovery rate across a variety of experimental conditions, enabling rapid and reliable genome-wide assessment of gene function and gene interactions in *S.cerevisiae*.

Introduction

Technologies that generate systematic genetic perturbations have revolutionized our ability to rapidly determine the genetic basis of diverse cellular phenotypes and behaviors (Burns et al., 1994; Deutschbauer et al., 2005; Giaever et al., 1999; Girgis, Liu, Ryu, & Tavazoie, 2007; Qian, Ma, Xiao, Wang, & Zhang, 2012; V. Smith, Botstein, & Brown, 1995; Winzeler et al., 1999). CRISPR-Cas9 technology (Cong et al., 2013; DiCarlo et al., 2013; Jinek et al., 2012), due to its high-fidelity and relatively low off-target effect, has become the dominant method for systematic, high-throughput genetic screening in diverse eukaryotic systems (Bassett, Kong, & Liu, 2015; Shalem et al., 2014; Sidik et al., 2016; Wang, Wei, Sabatini, & Lander, 2014). Nuclease-deactivated Cas9 (dCas9) has facilitated genome-wide screens further by enabling transient modulation of target gene expression (Bikard et al., 2013; Qi et al., 2013). This is achieved by fusing dCas9 to an inhibitory or activating domain to repress

(CRISPRi) or activate (CRISPRa) gene expression (Chavez et al., 2015; Dominguez, Lim, & Qi, 2015; Farzadfard, Perli, & Lu, 2013; Gilbert et al., 2013; Perez-Pinera et al., 2013; Smith et al., 2016). At the same time, the use of array-based oligonucleotide synthesis has enabled production of large-scale spacer libraries for use in genome-wide applications.

One complicating factor in the use of CRISPRi technology is balancing the efficacy of targeting with limiting the off-target activity of CRISPR/dCas9 machinery. Hence, many studies have aimed to determine the rules for efficient gRNA design (Farzadfard et al., 2013; Gilbert et al., 2013; Perez-Pinera et al., 2013; Qi et al., 2013). For example, in K562 human myeloid leukemia cells, optimal gRNAs are found to target a window of -50 bp to +300 bp relative to the transcription start site (TSS) of a gene (Gilbert et al., 2014). In *Saccharomyces cerevisiae,* however, the ideal guide positioning differs from human cell lines. Smith *et al.* (Smith et al., 2016) found that the optimal window is a 200 bp region immediately upstream of the TSS. The same group in a subsequent study refined their earlier findings showing that the region between TSS and 125 bp upstream of TSS is more effective for CRISPR-mediated repression (Smith et al., 2017). In addition, they showed a positive correlation between guide efficiency and chromatin accessibility scores, bolstering the notion that location of target in relation to TSS is not the only determinant of gRNA efficacy (Smith et al., 2016). The positioning and design rules of gRNAs are therefore organism specific. Recently, Lian et al (Lian, Schultz, Cao, HamediRad, & Zhao, 2019) developed a multi-functional genome-wide CRISPR (MAGIC) system for high throughput screening in *S.cerevisiae*. Although they successfully used a combinatorial approach to map the furfural resistance genes, their system did not utilize yeast-specific design rules (Smith et al., 2016). More importantly, their CRISPR system is not inducible making it difficult to perform context-dependent repression of gene function, and to survey the role of essential genes in arbitrary phenotypes of interest. Here, we introduce an inducible genome-scale library, dedicated for CRISPRi in *S. cerevisiae,* and designed based on the rules described by Smith *et al* (Smith et al., 2016; 2017). We demonstrate the efficacy of this library in targeting essential genes and identifying dosage-sensitive ORFs. In addition,

the library enabled us to identify genes involved in Adenine and Arginine biosynthesis using only a single round of selection. The CRISPRi library and protocol can, thus, be used to efficiently and inexpensively perform genome-wide knock-down screens to discover the genetic basis of a desired phenotype. In addition, the ability to easily perform CRISPRi screens in a desired genetic background of interest, enables rapid profiling of genetic interactions between a desired allele and knock-downs of all the genes in the genome.

Results and Discussion

Design and Construction of a Whole-Genome CRISPRi Library

We developed our CRISPRi library largely based on the design principles of Smith *et al*. (Smith et al., 2016) . Smith *et al*. (Smith et al., 2016) designed a single-plasmid inducible system expressing the catalytically inactive *Streptococcus pyogenes* Cas9 (dCas9) fused to the MXI1 transcriptional repressor (Gilbert et al., 2013), as well as a gRNA. The gRNA is under the control of a tetO-modified RPR1 RNA polymerase III promoter regulated by a tetracycline repressor (tetR), which is also expressed by this plasmid (Farzadfard et al., 2013; Hoon et al., 2008). Therefore, the expression of the gRNA is induced by addition of anhydrotetracycline (ATc) to the growth medium. In addition, a NotI site between the tetO and the gRNA scaffold enables the rapid cloning of spacers. TetR and dCas9-Mxi1 are expressed from the GPM1 and TEF1 promoters, respectively. For compatibility with an ongoing project in our group, we have replaced the *URA3* selection marker in PRS416 with *HIS3*. We call this plasmid amPL43 (Figure 1A). In order to validate the effectiveness of our system, we cloned gRNAs targeting *ERG25*, *ERG11* and *SEC14* genes in amPL43. Upon addition of ATc, these genes were repressed when compared to the samples without ATc (per qRT-PCR, Figure 1B). The repression was seen as early as one hour after induction and reached as much as 10-fold over 24 hours.

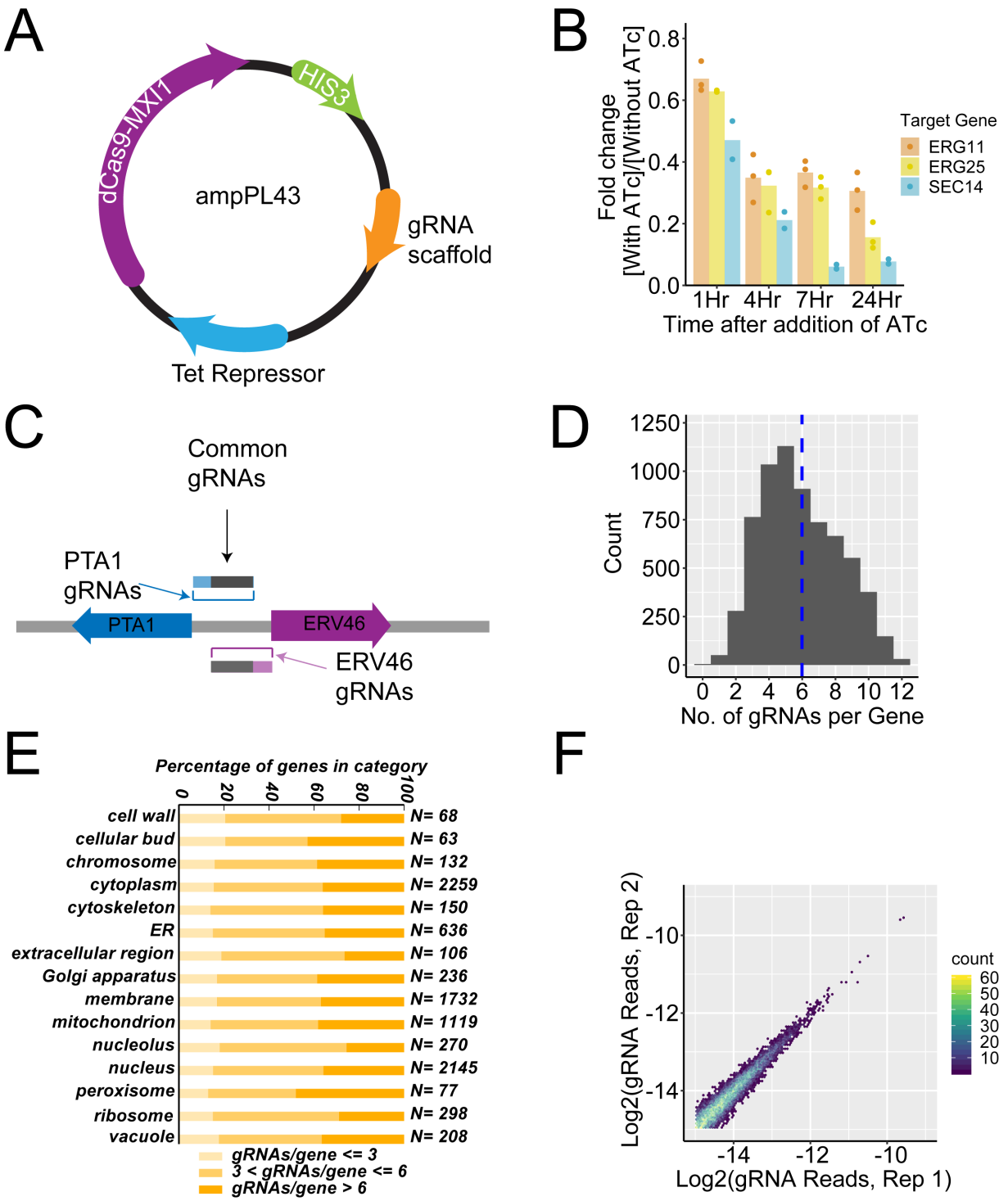

Figure 1. A) Schematic of amPL43 expression vector for inducible CRISPRi library in *S. cerevisiae*. gRNA scaffold region contains a NotI site for cloning gRNA sequences, an RPR1 promoter with a TetO site, and the constant region for the gRNA. **B)** The expression fold change of each target: ERG25 (three replicates), ERG11 (three replicates), and Sec14 (two replicates) as a result of gRNA induction by ATc, was calculated over time by qPCR. **C)** Schematic depicting the genomic region of PTA1 and ERV46. The gRNAs targeting the region between the two genes, depending on their proximity to each gene, could affect both genes. **D)** Histogram depicting the number of gRNAs per gene in replicate 1 of the library. The dashed blue line denotes the median ~6. **E)** The distribution of gRNAs per gene across different cellular compartments. **F)** Scatter plot depicting the frequency of reads per gRNA between two biological replicates of the yeast CRISPRi library (Spearman Correlation~0.98).

After confirming the effectiveness of this mode of CRISPR-inhibition, we constructed a genome-wide CRISPRi library to target all *S. cerevisiae* genes. To this end, we obtained and ranked all possible spacer sequences targeting every open reading frames (ORFs) using criteria described by Smith *et al*. (Smith et al., 2016; 2017). We then selected the top six gRNAs for each ORF. The dCAS9-MXI1 mediated repression could affect the genes on both plus and minus strands (Farzadfard et al., 2013). Therefore, some of the selected sequences could be targeting a neighboring gene. For example, a portion of the gRNAs targeting PTA1 could affect ERV46 (Figure 1C). For genes that have overlapping gRNA targets, we selected up to six additional sequences unique to those genes (See Methods). Altogether, we designed more than 52,000 gRNAs, with between 6-12 sequences per gene (Supplementary table 1). To generate a negative control set of gRNAs, we also included 500 randomly shuffled gRNAs with no matches in the yeast genome (Supplementary table 2).

This oligonucleotide library was synthesized on a 92918-format chip (CustomArray, Bothell, WA, USA) and cloned into amPL43 using universal adapter sequences. Briefly, the pooled oligonucleotides were amplified by PCR, cloned using Gibson Assembly (Gibson et al., 2009), and transformed into DH5α *E.coli* (NEB C2987H, >100x colonies/gRNA). The transformed bacteria were grown in a semisolid LB (0.35% Seaprep Agarose, Lonza, Basel, Switzerland) to minimize competition between the strains. Semisolid 3D media provides a more cost-effective and less labor-intensive method for large-scale libraries than conventional 2D plating. The pooled plasmid library was transformed into BY4741 using the standard LiAc/PEG method (Gietz & Schiestl, 2007) with minor modifications, and grown in semisolid SC-His+glu media (0.35% Seaprep Agarose, two biological replicates) for 48 hours, pooled and resuspended in SC-His+glu media, and frozen for future use.

For Next Generation Sequencing, the short gRNA sequences can act as unique identifiers of individual strains. Like barcodes, these can be quantified using next-generation amplicon sequencing. To this

end, the plasmids were extracted from yeast samples, and the targeting region was PCR amplified with flanking Illumina adapters and submitted for sequencing. More than 39,000 gRNAs were present in our CRISPRi library, targeting all but five of the yeast ORFs (YBL023C, YIL171W, YJL219W, YOR324C, YPR019W, Figures 1D and S1). There is a strong correlation between the frequency of reads of the gRNAs in the bacterial plasmid library and those transferred to yeast (Spearman Correlation~0.98). Two separate biological replicates of library transferred to yeast gave highly reproducible diversity and abundance (Spearman Correlation~0.98, Figures 1E-F, S2, and Supplementary Table 3).

High-Throughput Identification of Dosage sensitive Genes

To demonstrate the utility of our CRISPRi library for high-throughput genotype-phenotype mapping, we set out to determine whether simple outgrowth of the library could identify dosage sensitive genes. We first focused on identifying genes that have been determined to be haploinsufficient in yeast (Deutschbauer et al., 2005). One of the major advantages of our library is its inducibility. Inducible gRNA expression allows us to efficiently target dosage-sensitive genes for short intervals and determine their phenotypic outcome. To this end, we inoculated semisolid media with equivalent of 1 OD_{660} of the library (with an average of ~450 copies of each library member), with and without ATc induction. The fitness consequence of knocking down every yeast gene can be determined using the depletion or enrichment of the barcodes in the library in a comparison of induced and un-induced samples. Under these conditions, we expect that the gRNAs targeting haploinsufficient genes should be significantly depleted in the induced library (ATc+).

We grew the inoculated libraries for 24 hours, extracted plasmids from pooled samples, and prepared amplicons for NGS. We calculated the log-fold change of reads between samples with ATc and without ATc, for the two replicates (Fig 2A, Pearson correlation between two replicates ~0.65). Since the majority of the genes have more than three gRNAs targeting them, we sorted the gRNAs for each gene

based on their ratio and designated the mean of the most effective three gRNAs as the score for that gene (See Methods). Next, we averaged the scores for the two biological replicates. As expected, the gRNAs targeting haploinsufficient genes are significantly depleted in the induced compared to the uninduced sample (Figure 2B, *p*-value<$1x10^{-15}$ Mann-Whitney U-Test). To calculate the False Discovery Rate (FDR), we created a library of pseudogenes by sampling the scores from the random gRNAs to create a baseline (see methods for details). We used the receiver operating characteristic (ROC) curve analysis to find the threshold and to assign the categories and measure FDR. With only a single round of growth selection, and two replicates, we were able to correctly categorize ~79% of haploinsufficient genes with FDR<10% (Figures 2B-C and Supplementary Tables 4-5).

Next, we set out to determine what percentage of essential genes are dosage sensitive. The sensitivity and specificity of CRISPRi-based discovery of essential genes are limited by fundamental biological factors. On one hand, although essential genes are associated with functions that are indispensable to cellular life, it has been shown that not all essential genes are haploinsufficient and therefore dosage sensitive (Deutschbauer et al., 2005). This implies that even reducing the dosage of some essential genes to 50% would not measurably affect the fitness of the cell. On the other hand, essential genes are overrepresented among dosage-sensitive genes (Deutschbauer et al., 2005). Therefore, it is of interest to determine what percentage of essential genes show dosage-sensitivity and therefore can be detected by systematic CRISPRi knock-down. In fact, we observed that the majority of gRNAs targeting essential genes are significantly depleted, while the non-essential genes were largely unaffected (Figure 2B, *p*-value<$1x10^{-15}$ Mann-Whitney U-Test). Overall, we observed that ~59% of essential genes (FDR<10%) show significant gene depletion scores (Fig 2B, S3, *p*-value<$1x10^{-15}$ Mann-Whitney U-Test).

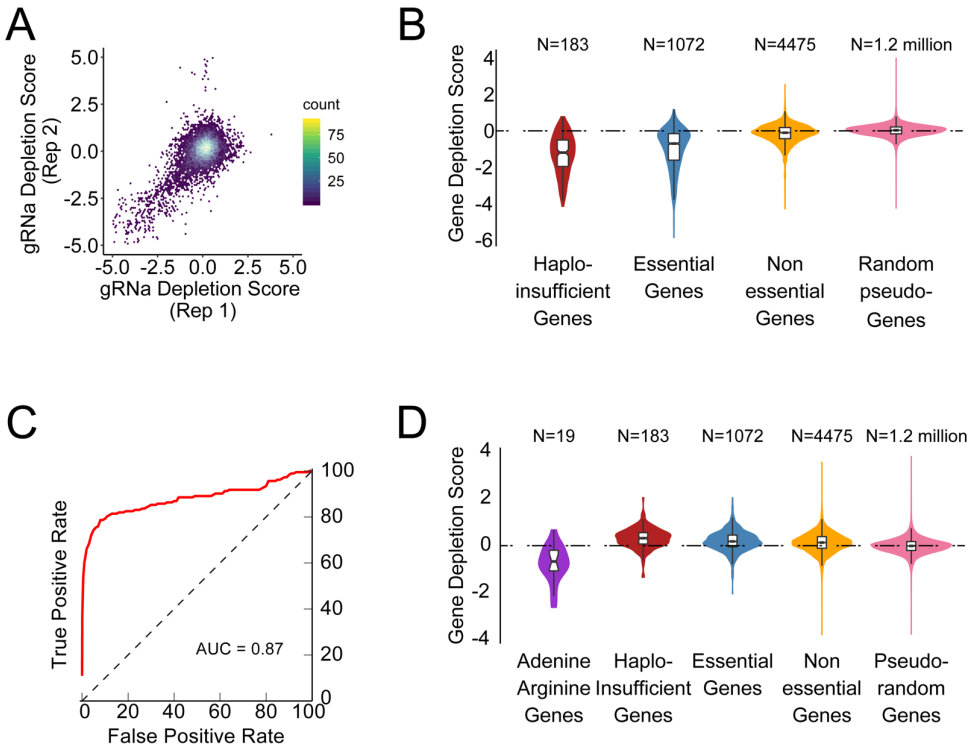

Figure 2. A) Binned scatter plot showing the gRNA depletion score between the two replicates (Pearson Correlation~0.65) **B)** Violin plots depicting the gene depletion score distribution for the haploinsufficient genes, essential genes, nonessential genes and random control pseudogenes, resulting from 200x random sampling of random gRNAs, in induced versus uninduced samples. **C)** ROC curve for the detection of haploinsufficient genes. Area under the curve is 0.87 **D)** Violin plots depicting the gene depletion score distribution for the adenine arginine deprivation experiment, shown for the known Adenine/Arginine Biosynthetic genes, haploinsufficient genes, essential genes, nonessential genes and random pseudogenes, resulting from 200x random sampling of random gRNAs.

The genome wide nature of our library, and its sensitivity to identify dosage-sensitive genes, gave us the opportunity to determine whether there was spatial bias for effective targeting by gRNAs. In particular, we were interested in factors such as, strand, distance to TSS, and nucleosome occupancy score. As can be seen in Figure S4A, in accordance with the previous finding by Smith *et al.* (Smith et al., 2017), gRNAs located between TSS and 125 bp upstream of TSS are particularly effective for detecting dosage sensitive genes. In addition, as may be expected, a higher nucleosome score seems to reduce the effectiveness of gRNA mediated transcriptional repression (Fig S4B). As expected, gRNAs for nonessential genes did not show any dependency between the gRNA depletion score and the distance to TSS or the nucleosome occupancy score (Fig S4C-D). Finally, we investigated whether gRNAs' effectiveness is influenced by the strandedness of gRNA targeting. To minimize the effect of

other factors, we only focused on gRNAs targeting dosage-sensitive essential genes that have a nucleosome occupancy score less than 0.1 and are within 125 bp upstream of TSS. The mean of the gRNA depletion score for gRNAs with the PAM on the same strand as the ORF was -1.53 while the mean for the gRNAs on the opposite strand was -1.02 (Fig S4E, Mann Whitney U-test *p*-value $<1x10^{-8}$). This indicates that, while gRNAs can target and repress ORFs on both strands, the strandedness significantly influences gRNAs' effectiveness.

Detection of the Genes Involved in Biosynthesis of Adenine and Arginine

Next, we explored whether our CRISPRi based approach can efficiently identify a smaller set of genes associated with a specific biological process. We chose to identify the genes involved in two distinct biosynthetic pathways, Arginine and Adenine, by comparing induced cells grown in drop-out media for histidine, arginine and adenine to cells grown in drop-out media for histidine. We expected gRNAs targeting genes that contribute to general cell fitness to be depleted to a similar degree for both samples, while gRNAs involved in the biosynthesis of arginine and adenine will be differentially affected. As such, we quantified the depletion of gRNAs in the arginine/adenine drop out media against the control. In order to systematically determine all the pathways that were affected, we used iPAGE (Goodarzi, Elemento, & Tavazoie, 2009), a sensitive pathway analysis tool that directly quantifies the mutual information between pathway membership and the distribution of gene depletion scores. The iPAGE results show that only the pathways for Arginine biosynthesis and purine biosynthetic processes (Crabeel, Soetens, De Rijcke, Pratiwi, & Pankiewicz, 1996; Jauniaux, Urrestarazu, bacteriology, 1978, n.d.; Jones & Fink, 1982; Ljungdahl & Daignan-Fornier, 2012; Messenguy, 1976; Myasnikov, Sasnauskas, Janulaitis, & Smirnov, 1991; Tibbetts & Appling, 1997) are significantly informative of the depletion of gRNAs targeting the pathways (Figures 2D, S5-6 and Supplementary Tables 4- 5, *p*-value$<1x10^{-7}$ Mann-Whitney U-Test). Our screen detected 7 out of 10 genes annotated in core arginine biosynthesis pathway (FDR<10%). Arginine biosynthesis is a particularly complex biosynthetic

pathway. It also has connections with several other pathways, such as polyamine and pyrimidine biosynthesis, and certain degradative pathways (Caldovic & Tuchman, 2003; Jones & Fink, 1982). Therefore, it is significant that we were able to detect its biosynthetic pathway with such precision. In addition, transcription of arginine biosynthetic genes is repressed in the presence of arginine by the ArgR/Mcm1p complex, which consists of Arg80p, Arg81p, Arg82p, and Mcm1p (Figure S6, Messenguy & Dubois, 2003). As such, downregulation of the ArgR/Mcm1p complex genes would be expected to increase the fitness of the cell in the absence of arginine in the media. Consistent with this expectation, we found that the gRNAs targeting these genes are enriched in the dropout media compared to the control media, providing further support for the sensitivity of our CRISPRi screen to detect positive and negative contributors to a phenotype of interest.

We also detected six out of nine genes involved in the annotated core adenine biosynthesis pathway (FDR<10%). Two of the three genes not detected, ADE16 and ADE17, are paralogs with partially redundant functions, and therefore would not be expected to be depleted in this screen (Tibbetts & Appling, 1997). Furthermore, although not in the canonical biosynthetic pathway, it is known that BAS1 and BAS2 activate the expression of adenine biosynthesis genes. Consistent with their positive contributions to adenine biosynthesis, our screen showed that gRNAs targeting BAS1 and BAS2 are also significantly depleted (Figures 2D, S6, Supplementary Tables 4-5).

It is important to point out that our FDR estimate of 10% is conservative, since strict gene ontology membership of these pathways does not capture the full complexity of the highly interactive genetic and regulatory networks coordinating nucleotide and amino-acid metabolism. For example, in addition to the genes involved in arginine and adenine biosynthesis, most depleted genes were involved in protein sorting and other general amino acid metabolic pathways such as SSY1 which is involved in sensing of external amino acid levels (Forsberg, Gilstring, Zargari, Martínez, & Ljungdahl, 2001). Finally, we looked more carefully at the genes whose interacting partners' gRNAs were depleted. Out

of these, VPS51, involved in protein sorting, had the highest enrichment for interactions with genes targeted by depleted gRNAs. This is of particular interest because mutations in many of the arginine genes could lead to abnormal vacuole morphology (Michaillat & Mayer, 2013). This provides further evidence that arginine deficiency undermines protein sorting functions of *S. cerevisiae*.

Conclusions

Here, we have established a powerful genome-wide functional screening library for CRISPRi in *S. cerevisiae*. Identification of dosage-sensitive genes provides a global validation for the general utility of this library. More importantly, our focused study of arginine/adenine deprivation demonstrates attractive sensitivity/specificity characteristics for probing the genetic basis of arbitrary phenotypes and biological processes. This CRISPRi library has distinct advantages: The gRNAs are specifically designed based on *S. cerevisiae* specific rules, and more importantly, the repression is inducible, enabling control over the scale, context, and timing of gene perturbations. The ability to quantitatively probe the role of essential genes is also a major advantage of inducible CRISPRi over both CRISPR and gene-deletion library approaches. Our use of 3D semisolid agarose to generate and interrogate large diverse libraries provides a more efficient approach over traditional 2D plating protocols while, at the same time, minimizing competitive biases that confound liquid outgrowth. Furthermore, the ability to easily transform the library into any genetic background of interest will enable rapid, parallel mapping of genetic interactions for any allele of interest (Jiang, Oikonomou, & Tavazoie, 2020). These advantages make our CRISPRi library a powerful and versatile tool for genetic interrogation of yeast biology.

Material and Methods

Strains, Plasmids and Media

PRS416-MXI1 was ordered from AddGene (Cambridge, MA, USA). In order to create amPL43, the PRS416-MXI1 backbone was PCR amplified. His3 was amplified from pSH62 (Gueldener, Heinisch, Koehler, Voss, & Hegemann, 2002). The primers are listed in Supplementary Table 6. Gibson assembly (NEB, Ipswich, MA, USA) was used to create amPL43. The plasmid's sequence was verified using Sanger sequencing. gRNA oligo library was purchased from Custom Array (92918-format chip, Bothell, WA, USA). The library spacers were amplified and extended as described by Smith *et al.* (Smith et al., 2016). To generate the library, amPL43 was maxiprepped (Qiagen, Hilden, Germany) and cut with NotI. Gibson Assembly was used to clone the oligos in the NotI site of amPL43. The library was transformed in DH5α *E.coli* (NEB C2987H). The transformed bacteria were grown in a 1L of semisolid LB plus 100 µg/ ml ampicillin (0.35% Seaprep Agarose, Lonza) to minimize competition between the colonies. To pool the library, the semisolid media was stirred for 10 minutes using a magnetic stirrer. 50 ml of the library was collected by centrifugation. The pooled library was miniprepped (Promega, Madison, WI, USA) according to manufacturer's instructions.

Competent BY4741 cells were produced using the standard LiAc/PEG method (Gietz & Schiestl, 2007) with minor modifications: 5mL of SC+glucose media was inoculated with a fresh colony of BY4741 and grown overnight in 30°C shaker. 1 ml of over-night culture (OD_{660}~ 6) was added to 25 mL of SC+glucose and shaken in 30°C shaker for 4 hours (OD_{660}~0.7). Cells were pelleted by centrifugation at 3000g for 5 minutes and washed twice in 25 mL sterile water. Next, cells were resuspended in 1 mL sterile water and transferred to 1.5 mL Eppendorf tube and centrifuged for 1 min at 4000g and then resuspended in sterile water to a final total volume of 1 mL. 200 µL of competent cells were aliquoted into an Eppendorf tube and centrifuged for 30 seconds at 4000g and the

supernatant was removed. Next, 240 µL of freshly made sterile 50 % PEG 3500, 36 µL of sterile 1 M lithium acetate, 50 µL of Salmon sperm DNA (Thermofisher, Waltham, MA, USA), and 2 µg of plasmid DNA was added to the cells and the volume was adjusted to 350 µL using sterile water. Then the mixture was vortexed and then incubated at 42° C for 20 minutes, vortexed, and again incubated at 42°C for another 20 minutes. Transformed cells were centrifuged and supernatant was removed and 1 mL of SC+glucose was added to it. The cells were moved to a round bottom falcon tube and shaken at 30°C for 1 hour.

The transformed library was grown in semisolid SC-His+glu media (0.35% Seaprep Agarose) for 48 hours and pooled and resuspended in SC-His+glu media and 20% glycerol and frozen for future use (two biological replicates). To extract the plasmids, 50 OD_{660} of cells were pooled and resuspended in 2 ml SE buffer (0.9M sorbitol, 0.1M EDTA pH 8.0), 20 µl Zymolyase 100T (2 mg/ml) and 20 µl beta mercaptoethanol and incubated at 37°C for one hour, followed by standard mini prep extraction per manufacturer's instructions (Promega). The pooled plasmid library was suspended in 50 µL of sterile water. For this study, we inoculated 250ml of the following three semisolid media with a number of cells equivalent to 1 OD660 unit of the library: A) SC-His with ATc, B) SC-His-Arg-Ade with ATc C) SC-His-Arg-Ade without ATc.

qPCRS

Strains were typically grown in SC-HIS overnight, diluted to an OD660 of 0.07 with/without 250 ng/mL ATc. We selected a mix of three different gRNAs per target gene. The cells were grown at 30°C and a sample was taken from each tube at 1, 4, 7 and 24hrs. RNA was extracted using Norgen Biotek Total RNA Purification Kit (Norgen Biotek, ON, Canada) and cDNA was made using Maxima H First Strand cDNA Synthesis Kit, with dsDNase (Thermofisher). ERG25, ERG11 and Sec14 primers and their corresponding spacers were adapted from Smith *et al.* (Smith et al., 2016). qRT-PCR was performed

using SYBR® Green PCR Master Mix (Life Technologies, Carlsbad, CA, USA) and the Quantstudio 5. ΔΔCt of the target genes' in induced versus uninduced states as compared to ACT1 level are reported.

Design of the gRNA Library

The gRNA sequences and their relative location to transcriptional start site (TSS), as well as the nucleosome occupancy scores, were adapted from Smith *et al.* (Smith et al., 2016). When possible, the gRNAs that were located between 0 and 125 bp of the TSS, were sorted based on their nucleosome occupancy score and the top six gRNAs were chosen. If six gRNAs were not available for any given gene, we searched for the gRNAs further from TSS. When more than six acceptable gRNAs were available, up to 12 gRNAs were included. For each gRNA shared between two genes, a unique gRNA was added, bringing the number of gRNAs per gene from six to 12 gRNAs per gene.

Next Generation Sequencing and Analysis

The extracted plasmids were amplified using "lib-seq" primers in the Supplementary Table 6 for 10 cycles using Phusion PCR kit (NEB). Each 50 µL PCR reaction consisted of 6 µL of plasmid library, 10 µL of 5X Phusion HF buffer, 0.5 µL of Phusion DNA Polymerase, 10 µM Forward Primer, 10 µM Reverse Primer, 10 mM dNTPs. Three replicate reactions per sample were amplified. The PCR conditions were: 98°C for 1 minute for initial denaturing, then 98°C for 18s, 66°C for 18s, and 72°C for 30s for 10 cycles then 72°C for 10 minutes. The products were purified using AMPure XP beads (Beckman-Coulter, Brea, CA, USA). The Illumina adapters and the indices were added by a second PCR using Q5 polymerase (NEB). Each 50 µL PCR consisted of 2-10 µL of purified PCR from the first PCR, 10 µL of 5X Q5 Reaction Buffer, 10 µL 5X Q5 High GC Enhancer, 0.5 µL of Q5 High-Fidelity DNA Polymerase, 10 µM Forward Primer, 10 µM Reverse Primer, 10 mM dNTPs. The PCRs were done in two steps. First step conditions were: 98°C for 1 minute for initial denaturing, then 98°C for 18s, 62°C for 18s, and 72°C for 30s for 4 cycles. Second stage of the PCR was 98°C for 18s,

66°C for 18s, and 72°C for 30s for variable cycles and finally 72°C for 10 minutes. The number of cycles should be determined using a parallel qPCR to ensure that the sample does not saturate. The samples were purified using AMPure XP beads. The samples were run on an Illumina Hi-Seq 4000 for 75 cycles paired-end with a 57-18 breakdown for read 1 and read 2. We used Cutadapt and Bowtie2 to trim the sequences and map them to the targeting sequences with a maximum of one mismatch (Langmead & Salzberg, 2012; Martin, 2011). We calculated the log frequency of the reads as:

$$log(f) = log_2\left(\frac{\text{No. of read}+1}{\text{Total number of reads for that sample + no. of gRNA species in the sample}}\right)$$

To calculate the depletion scores, we filtered the gRNAs to have a minimum read frequency of 10^{-5}. In addition, we excluded the gRNAs that were located further than 600 bp upstream of the TSS or if, in rare cases, they were present in multiple genomic locations. The gRNA depletion/enrichment score between sample A and B was calculated using:

$$gRNA\ depletion\ score = log(f_A) - log(f_B)$$

To assign a gene depletion or enrichment score, we sorted the gene log-fold change values of the gRNAs targeting a gene. Then we averaged the highest top three values and also bottom three values. We compared the absolute numbers for the top three and the bottom three and we assigned the biggest absolute value as the depletion/enrichment score for that gene. The gene scores for the two replicates were then averaged.

Haploinsufficient gene were derived from Deutschbauer *et al.,* 2005. Nonessential and essential gene lists were derived from Giaever *et al.,* 1999. To account for our experimental design, the haploinsufficient genes, and the genes involved in arginine, adenine or histidine biosynthesis were

removed from nonessential genes. To create the pseudogenes, we utilized the gRNA depletion scores of the random gRNAs. Then a pool of 200 times the number ORFs in the yeast genome was created. The distribution of gRNA per gene in this pool is the same as that of the observed library. The pseudogenes were populated by random sampling from the depletion scores of random gRNAs to create a library of random pseudogenes. To calculate the ROC curve, we oversampled the essential or haploinsufficient population to account for the imbalanced size of the positive and the negative groups.

IPAGE

We ran iPAGE (Goodarzi et al., 2009) in continuous mode on the average depletions/enrichment scores of guide RNAs calculated at the gene level as described above. iPAGE discovers gene categories that are significantly informative (p-value<0.001) of the average scores. Scores were sorted in 5 bins and only pathway categories were considered.

Acknowledgements

We thank members of the Tavazoie laboratory for helpful discussions and feedback on the manuscript, specially Wenyan Jiang and Balaji Santhanam. AMR was supported by a Ruth Kirschstein NRSA Postdoctoral Fellowship Award (F32-GM125170) from NIGMS. ST was supported by grants from the NIH (R01-AI077562 and R01-HG009065).

References:

Bassett, A. R., Kong, L., & Liu, J.-L. (2015). A genome-wide CRISPR library for high-throughput genetic screening in Drosophila cells. *Journal of Genetics and Genomics*, *42*(6), 301–309. http://doi.org/10.1016/j.jgg.2015.03.011

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., & Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research*, *41*(15), 7429–7437. http://doi.org/10.1093/nar/gkt520

Caldovic, L., & Tuchman, M. (2003). N-acetylglutamate and its changing role through evolution. *The Biochemical Journal*, *372*(Pt 2), 279–290. http://doi.org/10.1042/BJ20030002

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., P R Iyer, E., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. *Nature Methods*, *12*(4), 326–328. http://doi.org/10.1038/nmeth.3312

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819–823. http://doi.org/10.1126/science.1231143

Crabeel, M., Soetens, O., De Rijcke, M., Pratiwi, R., & Pankiewicz, R. (1996). The ARG11 gene of Saccharomyces cerevisiae encodes a mitochondrial integral membrane protein required for arginine biosynthesis. *Journal of Biological Chemistry*, *271*(40), 25011–25018. http://doi.org/10.1074/jbc.271.40.25011

Deutschbauer, A. M., Jaramillo, D. F., Proctor, M., Kumm, J., Hillenmeyer, M. E., Davis, R. W., et al. (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics*, *169*(4), 1915–1925. http://doi.org/10.1534/genetics.104.036871

DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Research*, *41*(7), 4336–4343. http://doi.org/10.1093/nar/gkt135

Dominguez, A. A., Lim, W. A., & Qi, L. S. (2015). Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nature Reviews Molecular Cell Biology*, *17*(1), 5–15. http://doi.org/10.1038/nrm.2015.2

Farzadfard, F., Perli, S. D., & Lu, T. K. (2013). Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synthetic Biology*, *2*(10), 604–613. http://doi.org/10.1021/sb400081r

Forsberg, H., Gilstring, C. F., Zargari, A., Martínez, P., & Ljungdahl, P. O. (2001). The role of the yeast plasma membrane SPS nutrient sensor in the metabolic response to extracellular amino acids. *Molecular Microbiology*, *42*(1), 215–228. http://doi.org/10.1046/j.1365-2958.2001.02627.x

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. http://doi.org/10.1038/nmeth.1318

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

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, *159*(3), 647–661. http://doi.org/10.1016/j.cell.2014.09.029

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, *154*(2), 442–451. http://doi.org/10.1016/j.cell.2013.06.044

Goodarzi, H., Elemento, O., & Tavazoie, S. (2009). Revealing Global Regulatory Perturbations across Human Cancers. *Molecular Cell*, *36*(5), 900–911. http://doi.org/10.1016/j.molcel.2009.11.016

Gueldener, U., Heinisch, J., Koehler, G. J., Voss, D., & Hegemann, J. H. (2002). A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Research*, *30*(6), e23–23. http://doi.org/10.1093/nar/30.6.e23

Hoon, S., Smith, A. M., Wallace, I. M., Suresh, S., Miranda, M., Fung, E., et al. (2008). An integrated platform of genomic assays reveals small-molecule bioactivities. *Nature Chemical Biology*, *4*(8), 498–506. http://doi.org/10.1038/nchembio.100

Jauniaux, J. C., Urrestarazu, L. A., bacteriology, J. W. J. O., 1978. (n.d.). Arginine metabolism in Saccharomyces cerevisiae: subcellular localization of the enzymes. *Am Soc Microbiol*.

Jiang, W., Oikonomou, P., & Tavazoie, S. (2020). *Comprehensive genome-wide perturbations via CRISPR adaptation reveal complex genetics of antibiotic sensitivity*. *Cell, in press*.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. http://doi.org/10.1126/science.1225829

Jones, E., & Fink, G. (1982). Regulation of amino acid and nucleotide biosynthesis in yeast. *The Molecular Biology of the Yeast Saccharomyces Metabolism and Gene Expression*, 181–299.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. http://doi.org/10.1038/nmeth.1923

Lian, J., Schultz, C., Cao, M., HamediRad, M., & Zhao, H. (2019). Multi-functional genome-wide CRISPR system for high throughput genotype-phenotype mapping. *Nature Communications*, *10*(1), 5794–10. http://doi.org/10.1038/s41467-019-13621-4

Ljungdahl, P. O., & Daignan-Fornier, B. (2012). Regulation of amino acid, nucleotide, and phosphate metabolism in Saccharomyces cerevisiae. *Genetics*, *190*(3), 885–929. http://doi.org/10.1534/genetics.111.133306

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. http://doi.org/10.14806/ej.17.1.200

Messenguy, F. (1976). Regulation of arginine biosynthesis in Saccharomyces cerevisiae: isolation of a cis-dominant, constitutive mutant for ornithine carbamoyltransferase synthesis. *Journal of Bacteriology*, *128*(1), 49–55.

Messenguy, F., & Dubois, E. (2003). Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene*, *316*, 1–21. http://doi.org/10.1016/s0378-1119(03)00747-9

Michaillat, L., & Mayer, A. (2013). Identification of genes affecting vacuole membrane fragmentation in Saccharomyces cerevisiae. *PloS One*, *8*(2), e54160. http://doi.org/10.1371/journal.pone.0054160

Myasnikov, A. N., Sasnauskas, K. V., Janulaitis, A. A., & Smirnov, M. N. (1991). The Saccharomyces cerevisiae ADE1 gene: structure, overexpression and possible regulation by general amino acid control. *Gene*, *109*(1), 143–147. http://doi.org/10.1016/0378-1119(91)90600-g

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods*, *10*(10), 973–976. http://doi.org/10.1038/nmeth.2600

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173–1183. http://doi.org/10.1016/j.cell.2013.02.022

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelson, T., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, *343*(6166), 84–87. http://doi.org/10.1126/science.1247005

Sidik, S. M., Huet, D., Ganesan, S. M., Huynh, M.-H., Wang, T., Nasamu, A. S., et al. (2016). A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. *Cell*, *166*(6), 1423–1435.e12. http://doi.org/10.1016/j.cell.2016.08.019

Smith, J. D., Schlecht, U., Xu, W., Suresh, S., Horecka, J., Proctor, M. J., et al. (2017). A method for high-throughput production of sequence-verified DNAlibraries and strain collections. *Molecular Systems Biology*, *13*(2), 913–15. http://doi.org/10.15252/msb.20167233

Smith, J. D., Suresh, S., Schlecht, U., Wu, M., Wagih, O., Peltz, G., et al. (2016). Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome Biol*, 1–16. http://doi.org/10.1186/s13059-016-0900-9

Tibbetts, A. S., & Appling, D. R. (1997). Saccharomyces cerevisiae expresses two genes encoding isozymes of 5-aminoimidazole-4-carboxamide ribonucleotide transformylase. *Archives of Biochemistry and Biophysics*, *340*(2), 195–200. http://doi.org/10.1006/abbi.1997.9919

Wang, T., Wei, J. J., Sabatini, D. M., & Lander, E. S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, *343*(6166), 80–84. http://doi.org/10.1126/science.1246981

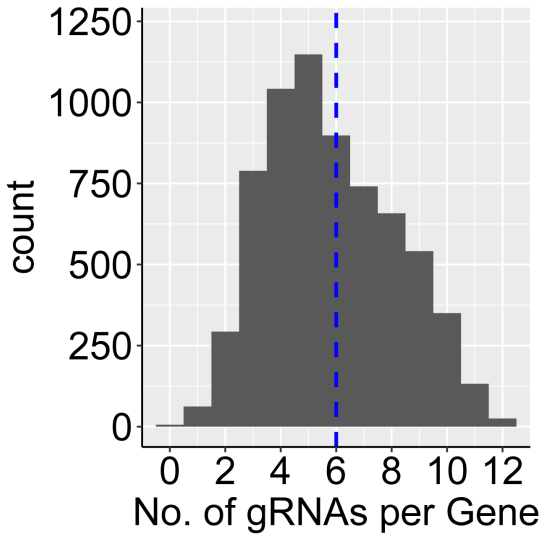

Figure S1. Histogram depicting the number of gRNAs per gene in library cloned in bacteria. The dashed blue line denotes the median: ~6.

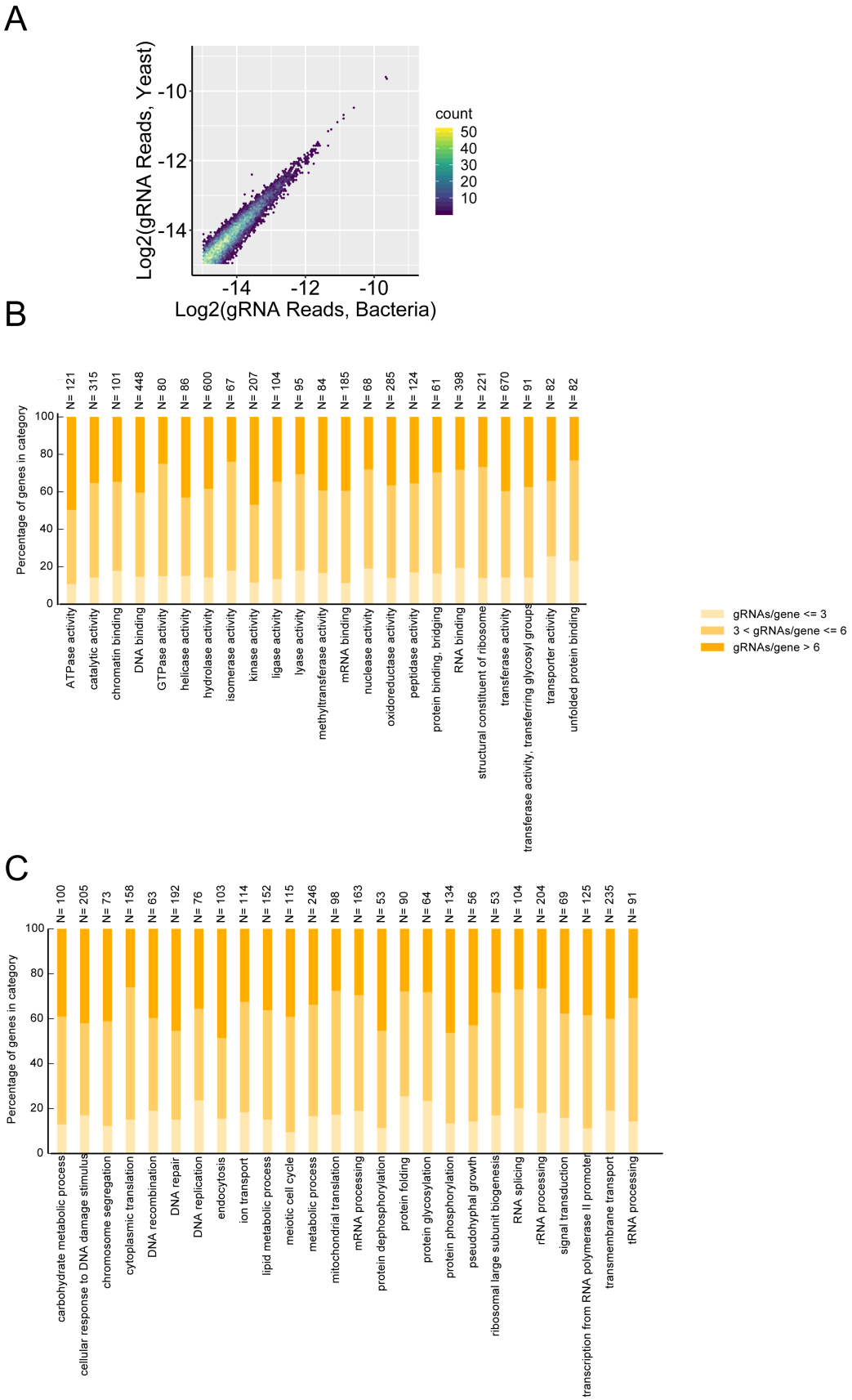

Figure S2. A) Scatter plot depicting the frequency of reads per gRNA between library cloned in bacteria and yeast biological replicate 1, showing a strong correlation (Spearman Correlation~0.98). **B)** The distribution of gRNAs per gene across different molecular functions. **C)** The distribution of gRNAs per gene across different biological processes.

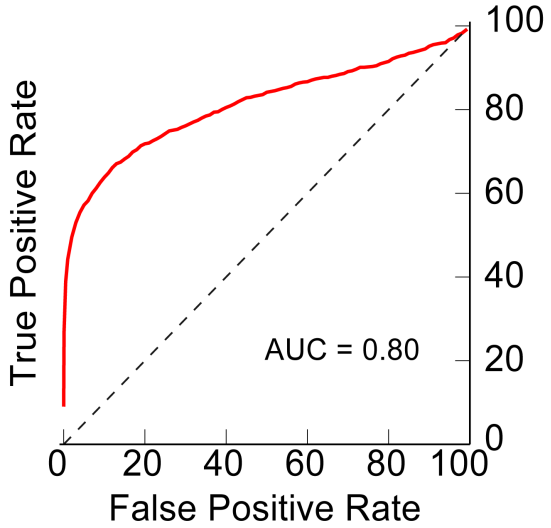

Figure S3. ROC curve for the detection of essential genes. Area under the curve is 0.80.

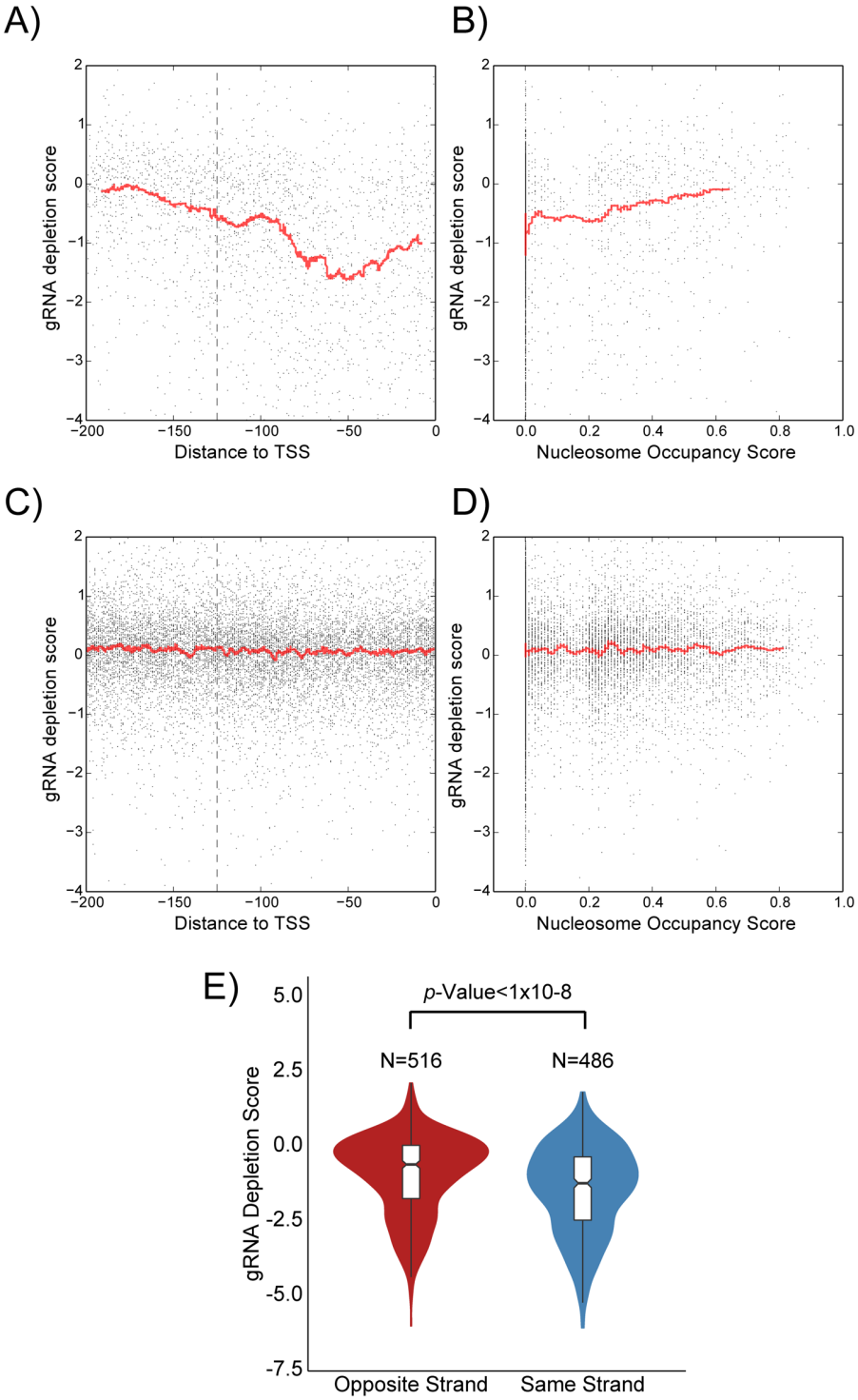

Figure S4. A) Scatter plot depicting the gRNA depletion score for dosage sensitive essential genes versus the distance from TSS. The red line represents the rolling average (window of 200). The dashed line signifies 125 bp upstream of TSS. **B)** Scatter plot depicting the gRNA depletion score for dosage sensitive essential genes versus the nucleosome occupancy score. The red line represents the rolling average (window of 200). **C)** Scatter plot depicting the gRNA depletion score for nonessential genes versus the distance from TSS. The red line represents the rolling average as above. The dashed line signifies 125 bp upstream of TSS. The red line represents the rolling average as above. **D)** Scatter plot depicting the gRNA depletion score for nonessential genes versus the nucleosome occupancy score. The red line represents the rolling average as above. E) Violin plots depicting the distribution of gRNA depletion scores for gRNAs targeting the opposite or same strand as the target ORF.

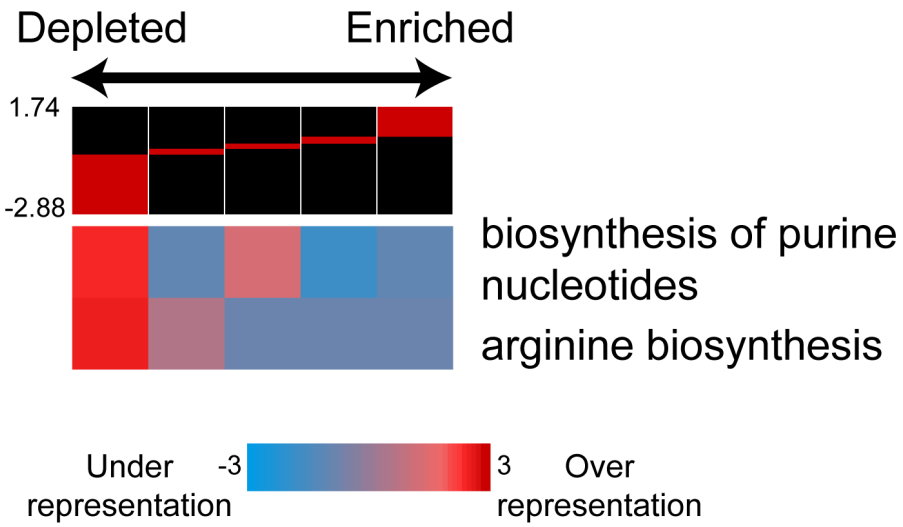

Figure S5. iPAGE results shows all the biological pathways that are perturbed in a statistically significant manner as a result of adenine arginine starvation.

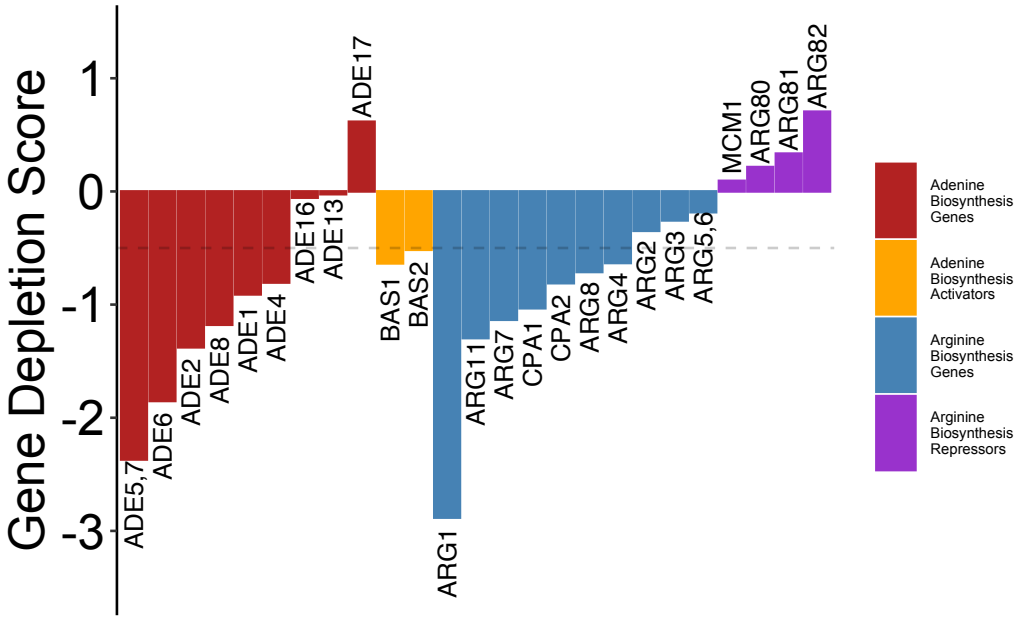

Figure S6. Bar plots showing the gene depletion scores for the genes involved in adenine and arginine biosynthesis pathways in addition to their regulatory genes. Dashed line signifies the threshold for FDR= 10%.

Supplementary Tables:

Table 1. Summary of the designed spacers. When possible, the gRNAs that were located between 0 and 125 bp of the TSS, were sorted based on their nucleosome occupancy score and the top six gRNAs were chosen. If six gRNAs were not available for any given gene, we identified additional gRNAs further from TSS.

Table 2. Summary of the control random spacers.

Table 3. Summary of frequency of reads in each sample.

Table 4. Sheet (1) Gene depletion scores for samples with ATc versus without ATc. **Sheet (2)** Gene depletion scores for samples in the drop-out media for Histidine, Arginine and Adenine versus Histidine drop-out media

Table 5. List of the gene categories used in this study

Table 6. List of the primer sequences used in this study