

Chapter 7

Chemogenomic Approaches to Elucidation of Gene Function and Genetic Pathways

Sarah E. Pierce, Ronald W. Davis, Corey Nislow, and Guri Giaever

Summary

The ~6,000 strains in the yeast deletion collection can be studied in a single culture by using a microarray to detect the 20 bp DNA “barcodes” or “tags” contained in each strain. Barcode intensities measured by microarray are compared across time-points or across conditions to analyze the relative fitness of each strain. The development of this pooled fitness assay has greatly facilitated the functional annotation of the yeast genome by making genome-wide gene-deletion studies faster and easier, and has led to the development of high throughput methods for studying drug action in yeast. Pooled screens can be used for identifying gene functions, measuring the functional relatedness of gene pairs to group genes into pathways, identifying drug targets, and determining a drug’s mechanism of action. This process involves five main steps: preparing aliquots of pooled cells, pooled growth, isolation of genomic DNA and PCR amplification of the barcodes, array hybridization, and data analysis. In addition to yeast fitness applications, the general method of studying pooled samples with barcode arrays can also be adapted for use with other types of samples, such as mutant collections in other organisms, siRNA vectors, and molecular inversion probes.

Key words: *S. cerevisiae*, Drug-target identification, Functional assays, Chemogenomics, Gene networks, Genomics, DNA barcodes

1. Introduction

The yeast deletion collection is a set of single-gene deletion strains in *S. cerevisiae* that covers almost every open reading frame (ORF) in the genome (96%) (1). The deletion collection was created to provide a pre-constructed, saturated set of gene-deletion mutants that can be used in the place of random-mutant libraries and individually constructed strains. In addition to fulfilling this

goal, the collection has also enabled the development of methods for studying all ~6,000 deletion strains in a single culture (1–3). These methods are made possible by the inclusion of unique 20 bp DNA “barcodes” or “tags” in each strain (Fig. 1). that enables relative strain concentrations to be compared by hybridizing pooled samples to a microarray that contains the tag complements (Fig. 2). The development of this pooled fitness assay has allowed genome-wide screens to be performed in thousands of conditions, providing a wealth of functional information about the yeast genome (1–15).

Pooled fitness assays have a variety of uses. Gene function can be studied by screening the collection in a condition whose impact on the cell is already known (Fig. 3a). For example, if a strain deleted for gene X is sensitive to a DNA damaging agent, it suggests that gene X is involved in DNA damage response (9,15, 18). This is a stronger test of gene function than examining changes in conditional expression because the phenotypic effect of each gene deletion is measured directly. This claim is supported by the observation that genes required for growth in a particular condition are often not overexpressed in that condition, and that a change in expression does not always indicate functional involvement (1, 4, 15, 16, 19).

Pooled assays can also be used to study drug function. The function of a drug can often be determined by identifying the gene deletions that confer increased drug sensitivity or resistance

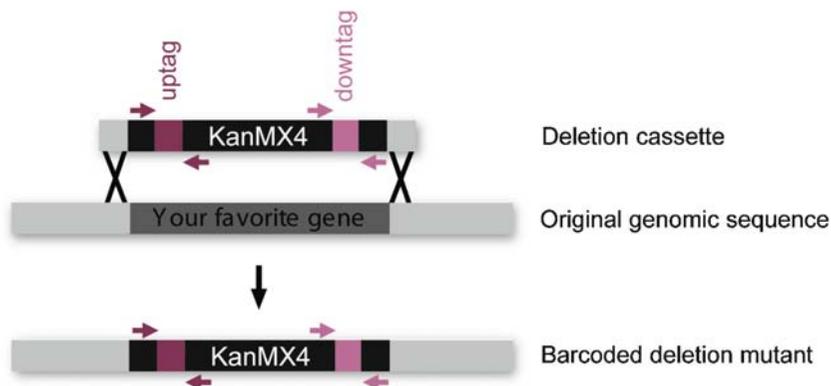


Fig. 1. Deletion cassette used for constructing strains in the yeast deletion collection. Each cassette carries the G418-resistance marker KanMX4, which is required for the selection of transformants. The marker is flanked by two unique “barcode” or “tag” sequences, which are called the “uptag” and the “downtag.” These barcodes were designed to be maximally distinct, have uniform hybridization properties, and to carry no homology to the yeast genome (see Note 12). The two barcode sequences are flanked by four universal primer sites (arrows) that are common to all strains, and allow the barcodes to be amplified from a pooled culture. Uptags and downtags are amplified separately to avoid primer cross-talk. Integration into the genome is directed by two 45-bp homology regions which mediate replacement of the targeted gene by mitotic recombination (shown as *black cross marks*). Modified from ref. 38.

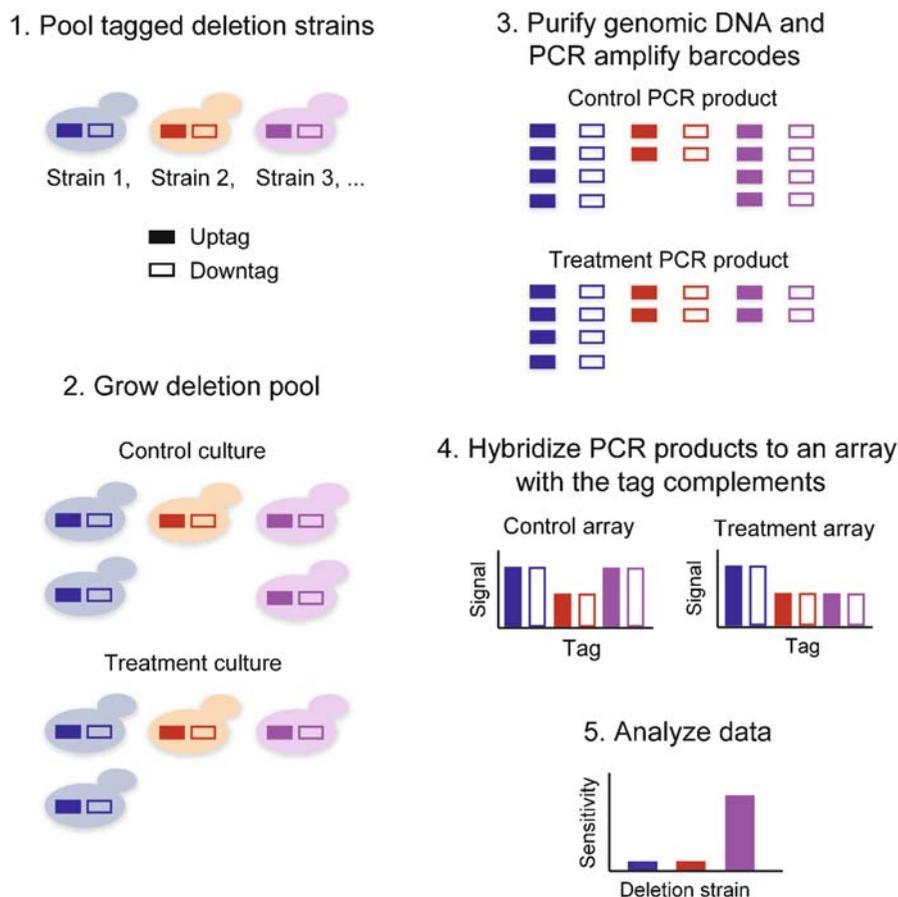


Fig. 2. Overview of the pooled fitness assay. Fitness profiling of pooled deletion strains involves five main steps: (1) Strains are first pooled at approximately equal abundance. (2) The pool is grown competitively in the condition of choice and a control condition. If a gene is sensitive to the treatment condition, the strain carrying this deletion will grow more slowly and become under-represented relative to the control culture (strain 3). Resistant strains will grow faster and become over-represented (not shown). (3) Genomic DNA is isolated from cells harvested at the end of pooled growth, and barcodes are amplified from the genomic DNA with universal primers. (4) PCR products are then hybridized to an array that detects the tag sequences, giving tag intensities for the two samples. (5) The treatment and control sample are then compared to determine the relative fitness of each strain. Note that only strain 3 is called as sensitive to the condition. While strain 2 grows more slowly than strain 1 in the treatment, this growth difference is not of interest because it matches that seen in the control. Modified from ref. 38.

because many genes in the yeast genome have annotated functions, making it likely that any given drug will perturb deletion strains with existing functional information (**Fig. 3b**). For example, if drug X perturbs deletion strains known to be sensitive to DNA damage, it suggests that drug X is a DNA damaging agent (**13**). Sensitivity data can also be used to identify subtle differences between compounds with similar modes of action, as well as secondary effects that may remain unknown even if the drug's primary function is well characterized (**9, 13, 16**).

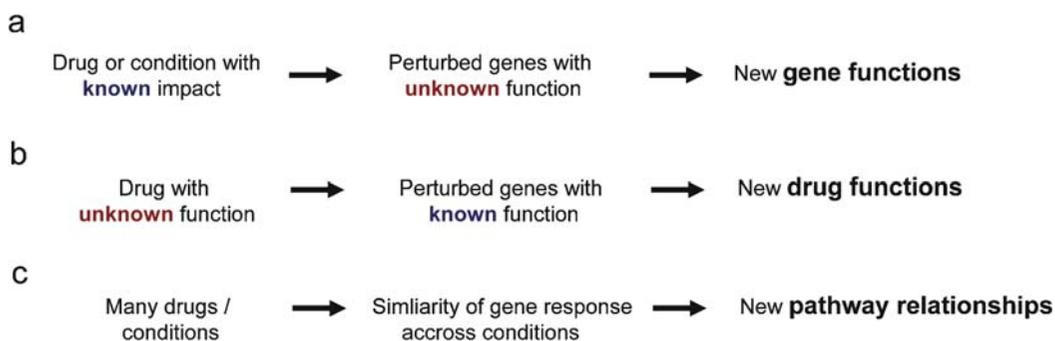


Fig. 3. Applications of homozygous deletion screening. Homozygous deletion screens can be used for studying gene function, drug function, and gene relationships. **(a)** Screening the homozygous deletion collection in a condition with a known impact on the cell provides functional information about the sensitive and resistant genes identified. **(b)** Screening a drug of unknown function will typically perturb some genes with existing functional annotations. These identified genes give information about the drug's mode of action. **(c)** Genes that are functionally related will typically be sensitive in similar conditions. The combined results of many homozygous screens can therefore be used to give information about the functional relatedness of gene pairs. Functional relatedness is estimated by measuring the similarity of a gene-deletion's response across many conditions.

In addition to generating information about single genes and single drugs, large sets of experiments can be used to study the relationships between genes and the relationships between drugs (**Fig. 3c**). For a given gene pair, the correlation of fitness values across conditions can be used to estimate the functional relatedness of the two genes, enabling large functional datasets to identify functional relationships between genes. Similarly, for a given drug pair the correlation of fitness values across genes is a measure of the similarity of their mechanism of action (15).

Pooled analysis of the heterozygous deletion strains can be used to identify novel drug targets (**Fig. 4**) (6, 7, 10, 17). The reasoning behind this technique is as follows: if deletion of the drug target results in reduced growth, then heterozygous deletion of the gene encoding this target will often confer increased sensitivity to the drug. This is because the gene copy number can affect protein abundance (20), and if the heterozygote has a reduced level of the target protein, a lower dose of drug will be required to deplete enough of the protein's activity to mimic the phenotype of the gene's homozygous deletion, which is reduced growth. The heterozygote will therefore require a lower dose of drug than the wild-type strain to show the slow-growth phenotype that result from full inactivation of the target protein. This effect enables all genes whose full deletion confers growth sensitivity to be screened for possible drug targets in a single pooled experiment.

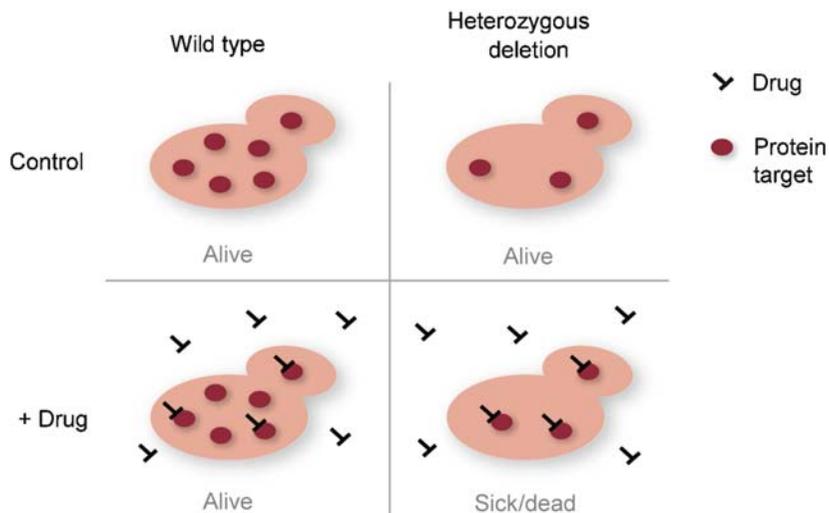


Fig. 4. Heterozygous profiling for drug target identification. Screening the heterozygous deletion collection can identify drug targets for genes whose homozygous deletion causes a growth defect. This method exploits the fact that a gene's copy number is related to its expression level. Here a drug with a specific protein target is shown in *black*, and the corresponding protein target is shown as a *circle*. At an intermediate concentration, the drug will inactivate most target function in the heterozygote, but not in strains with two functional gene copies. This will result in a growth defect for the strain that carries a heterozygous deletion of the gene encoding the drug's target. This phenotype mimics the growth defect caused by the corresponding homozygous deletion. Modified from ref. 38.

In addition to the applications mentioned above, these protocols can also be used to adapt the DNA barcode approach to other types of samples. Barcoded sample-tracking is a robust method that can be applied to a variety sample types. For example, DNA barcodes have already been used to tag mutant collections in other organisms (21, 22), collections of siRNA vectors (23–30), and DNA probes such as molecular inversion probes (MIPs) (31–34).

The pooled fitness assay involves five main steps (Fig. 2, for an experimental timeline see Fig. 5). First, the deletion collection is grown on solid media in an arrayed format and the resulting cells are pooled and frozen to make cell aliquots that will be used for starting growth experiments. Cells are then grown in the desired conditions, typically for 5–20 generations. The barcodes from the resulting cell samples are prepared for hybridization by isolation of the genomic DNA and PCR amplification with the common barcode primers. Amplification is performed in two separate reactions to prevent crosstalk between the uptag and dntag primers. The resulting PCR products are then hybridized to tag microarrays. The barcode intensity data is then analyzed to determine differences in strain representation between pairs of samples.

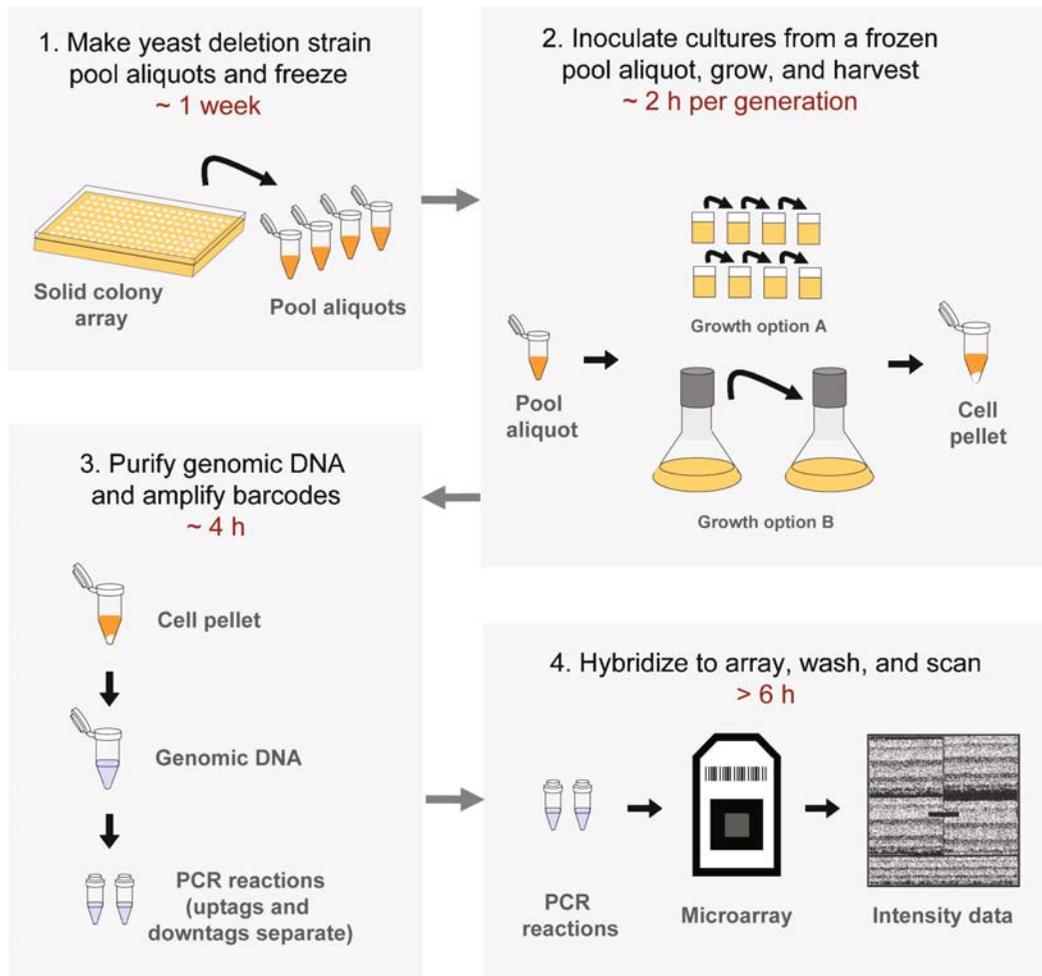


Fig. 5. Timeline for a pooled growth experiment. (1) Pooled cells are generated by transferring a copy of the deletion collection from solid media to liquid and then freezing small aliquots. This step is time consuming, but is performed infrequently because many aliquots can be generated at once. (2) Cultures are inoculated using a frozen cell aliquot, and then grown for the desired number of generations. The specific amount of time needed for growth will vary depending on the number of generations, and on the growth rate of the culture in the treatment condition. The example shown is for 20 generations of growth either as (a) 700- μ l cultures in a 48-well plate, shaken and grown in a spectrophotometer, or (b) 50-ml cultures grown in flasks. (3) Genomic DNA is purified from the harvested cells using a standard column-based purification kit, and tags are then PCR amplified from the purified genomic DNA. The uptags and downtags are amplified separately to avoid cross-reactions between the uptag and downtag primer pairs. (4) PCR products from both reactions are hybridized to a single array. After hybridization, the array is washed and scanned. The time needed for this step will vary depending on the hybridization time chosen. Modified from ref. 38.

2. Materials

2.1. Preparing Pooled Cell Aliquots

1. Frozen glycerol stocks of the yeast deletion collection in 96-well microtiter plates (OpenBiosystems, Part Nos. YSC1056 and YSC1055).

2. Nunc Omni Trays (VWR, Catalog No. 62409-600).
3. 96-Well pin tool (V&P Scientific, Catalog No. VP407A).
4. EtOH for flame-sterilizing pin tool (Gold Shield, Catalog No. DSP-CA-151).
5. Incubator set to 30°C for growing yeast plates.
6. Spectrophotometer.
7. 1,000× G418 Stock (200 mg/ml): Dissolve 5 g G418 (Agri-Bio, Catalog No. 3000) in 25 ml dH₂O. Filter sterilize using a 0.2- μ m filter and a syringe. Shield from light by wrapping bottle in foil. Store at 4°C.
8. YPD + 200 μ g/ml G418 rectangular plates: Mix 10 g yeast extract (Becton Dickinson and Co., Catalog No. 212750), 20 g peptone (Becton Dickinson and Co., Catalog No. 211677), 20 g dextrose Becton Dickinson and Co., Catalog No. 215530), 20 g agar (Becton Dickinson and Co., Catalog No. 214010), and 1 l dH₂O to a 2-l flask with a stir bar. Autoclave. Allow media to cool to approximately 50°C with gentle stirring. Add 1 ml 1,000× G418 Stock. Stir gently for an additional 1 min to ensure drug is evenly mixed. Pour into Nunc Omni trays, 50 ml per tray. Sufficient for approximately 20 plates. Store at 4°C.
9. YPD liquid + 200 μ g/ml G418: Mix 10 g yeast extract (Becton Dickinson and Co., Catalog No. 212750), 20 g peptone (Becton Dickinson and Co., Catalog No. 211677), 20 g dextrose Becton Dickinson and Co., Catalog No. 215530), and 1 l dH₂O to a 1-l bottle. Autoclave. Allow media to cool to approximately 50°C. Add 1 ml 1,000× G418 Stock. Store at 4°C.

2.2. Growing Pooled Cultures

1. Temperature controlled shaker for 250-ml flasks or spectrophotometer that allows incubation and shaking of plates, such as the TECAN GENIOS SpectraFLUOR Plus (Tecan, Part No. F129005, Note: many spectrophotometers do not shake plates hard enough for growing yeast).
2. 250-ml culture flasks (if growing cultures in flasks).
3. 48-Well plates (Greiner, Part No. 677102; if growing cultures in plates).
4. Plate roller (Sigma, Catalog No. R1275) for sealing 48-well culture plates (if growing cultures in 48-well plates).
5. Adhesive plate seals (ABgene, Catalog No. AB-0580) for sealing 48-well culture plates (if growing cultures in 48-well plates).
6. Desired media and compounds for growing cultures.

2.3. Purification and Amplification of Barcodes from Cell Samples

1. YeaStar Genomic DNA Kit (Zymo Research, Catalog No. D2002).
2. Taq DNA polymerase (Invitrogen, Catalog No. 10342).
3. dNTPs (Invitrogen, Catalog No. 10297).
4. 10× PCR reaction buffer (MgCl₂) (Invitrogen, Part No. 52724).
5. 50 mM MgCl₂ (Invitrogen, Part No. 52723).
6. Up primer mix: Dissolve Uptag (5′ – GAT GTC CAC GAG GTC TCT – 3′) and Buptagkanmx4 (5′ biotin – GTC GAC CTG CAG CGT ACG – 3′) each in dH₂O at 100 pmol/μl, then mix in a 1:1 ratio for a final concentration of 50 pmol/μl each. Store at –20°C.
7. Down primer mix: Dissolve Dntag (5′ – CGG TGT CGG TCT CGT AG – 3′) and Bdntagkanmx4 (5′ biotin – GAA AAC GAG CTC GAA TTC ATC G – 3′) each in dH₂O at 100 pmol/μl, then mix in a 1:1 ratio for a final concentration of 50 pmol/μl each. Store at –20°C.
8. Thermocycler with heated lid.

2.4. Array Hybridization and Scanning

1. Genflex Tag 16K Array v2 (Affymetrix, Part No. 511331).
2. Hybridization Oven 640 (Affymetrix, Part No. 800138).
3. GeneChip Fluidic Station 450 (Affymetrix, Part No. 00-0079).
4. GeneArray Scanner 3000 (Affymetrix, Part No. 00-0212).
5. 0.5-ml microfuge tubes suitable for boiling (Eppendorf 0.5-ml Safe-Lock microcentrifuge tubes, Sigma, Catalog No. T8911).
6. Microcentrifuge tube holders (Sigma, Catalog No. Z708372).
7. Ice bucket.
8. Boiling water bath with floating rack for 0.5-ml tubes.
9. Teeny Tough-Spots (Diversified Biotech, Catalog No. TS-TNY; or any other small labels suitable for preventing evaporation from Affymetrix array gaskets).
10. Denhardt's Solution, 50× Concentrate (e.g., Sigma, Catalog No. D-2532).
11. Streptavidin, R-phycoerythrin conjugate (SAPE) – 1 mg/ml (Invitrogen, Catalog No. S-866). Store at 4°C. Do not freeze.
12. B213 oligonucleotide: Dissolve B213 control (5′ biotin – CTGAACGGTAGCATCTTGAC – 3′) in dH₂O at 100 pm/μl to make a concentrated stock, and then use the stock to make 1 fmol/μl working solution. Store at –20°C.

13. Mixed oligonucleotides: Dissolve each of the following eight oligos in dH₂O at 100 pmol/μl: Uptag (5' – GAT GTC CAC GAG GTC TCT – 3'), Dntag (5' – CGG TGT CGG TCT CGT AG – 3'), Uptagkanmx (5' – GTC GAC CTG CAG CGT ACG – 3'), Dntagkanmx (5' – GAA AAC GAG CTC GAA TTC ATC G – 3'), Uptagcomp (5' – CTA CAG GTG CTC CAG AGA – 3'), Dntagcomp (5' – GCC ACA GCC AGA GCA TC – 3'), Upkancomp (5' – CAG CTG GAC GTC GCA TGC – 3'), Dnkancomp (5' – CTT TTG CTC GAG CTT AAG TAG C – 3'). Mix an equal volume of each of the eight oligos for a final concentration of 12.5 pmol/μl each. Note that the Uptag oligo is also used in Up primer mix and the Dntag oligo is also used in Down primer mix, so take care to leave enough for use in both mixes. Store at –20°C. 5 M NaCl Dissolve 58.4 g NaCl in 200 ml dH₂O. Store at room temperature (18–25°C).
14. 0.5 M EDTA: Dissolve 7.3 g EDTA (BioRad, Catalog No. 161-0729) in 50 ml dH₂O. Store at room temperature.
15. 10% Tween: Mix 2 ml Tween 20 (Sigma, Catalog No. T2700) with 18 ml dH₂O. Store at 4°C.
16. 1% Tween: Mix 10 μl Tween 20 (Sigma, Catalog No. T2700) with 990 μl dH₂O. Store at 4°C.
17. 12× MES stock: Mix 0.70 g MES free acid monohydrate (Sigma, Catalog No. M5287), 1.9 g MES Sodium Salt (Sigma, Catalog No. M5057), 8 ml of H₂O and adjust volume to 10 ml. Adjust the pH to be between 6.5 and 6.7 using a pH meter or pH paper. Shield from light by wrapping bottle in foil. Store at 4°C. Replace if solution becomes visibly yellow or after 1 month.
18. 2× Hybridization buffer: Mix 8.3 ml of 12× MES Stock, 17.7 ml of 5 M NaCl (J.T. Baker, Catalog No. 3624-01), 4.0 ml of 0.5 M EDTA, 0.1 ml of 10% Tween 20 (vol/vol), 19.9 ml filtered dH₂O. Filter to remove dust particles that would interfere with scanning. Shield from light by wrapping bottle in foil. Store at 4°C. Replace as for MES stock.
19. Wash A: Mix 300 ml 20× SSPE (Sigma, Catalog No. S2015), 1 ml 10% Tween (vol/vol), 699 ml filtered dH₂O. Filter to remove dust particles that would interfere with scanning. Store at room temperature.
20. Wash B: Mix 150 ml 20× SSPE (Sigma, Catalog No. S2015), 1 ml 10% Tween (vol/vol), 849 ml dH₂O. Filter to remove dust particles that would interfere with scanning. Store at room temperature.

3. Methods

3.1. Yeast Deletion

Strain Pool

Construction

1. Allow at least 1 week to generate pooled aliquots of cells before beginning pooled growth experiments. Pooling is performed infrequently because pooled cells can be stored indefinitely at -80°C . See **Note 1** for information about strain background options.
2. Allow the frozen glycerol stocks for the strains of interest to thaw completely because cells may have settled prior to being frozen. Remove plates from the freezer in manageable numbers to avoid leaving all the plates at room temperature for an extended period.
3. To sterilize a 96-well pin tool, dip pin tool in water to rinse away any remaining cells, followed by two dips in 70% ethanol baths, then carefully flame the pin tool. Allow pin tool to cool for 1 min. Make certain that the level of the ethanol bath exceeds the level in the water bath to ensure all carry-over cells are flamed and removed. Change water frequently.
4. Insert the sterile 96-well pin tool into a thawed 96-well plate, swirl gently to capture settled cells and then transfer to a Nunc Omni Tray containing 50 ml of YPD-agar including 200 $\mu\text{g}/\text{ml}$ G418. Allow pin to dwell on agar for 5–10 s. Repeat for the remaining plates, sterilizing the pin tool between transfers.
5. Grow colonies until they reach maximal size at 30°C (2–3 days).
6. After colonies have reached full size, make note of any strains that are missing or appear as slow-growing colonies. Streak these strains out individually using standard yeast procedures to generate more cells.
7. Scrape the entire contents of all plates (in a laminar flow hood to avoid contamination) into a 50-ml conical centrifuge tube containing YPD liquid media + 200 $\mu\text{g}/\text{ml}$ G418. Alternately, cells can be soaked off the plates by adding 12 ml YPD liquid media + 200 $\mu\text{g}/\text{ml}$ G418 to each plate, soaking for 5 min, and gently agitating the plate to resuspend the cells. After cells are resuspended, pool liquid in a flask by transferring a fixed volume from each plate with a pipette.
8. For the slow-growing strains observed in **step 5**, add approximately three colony equivalents of cells using a sterile flat toothpick. This will help to ensure that these slow growing strains remain detectable by the end of a typical growth experiment.
9. Add glycerol to 15% or DMSO to 7% (vol/vol).

10. Measure the OD_{600} of the pool and adjust to a final concentration of 50 OD_{600}/ml (OD_{600} 1.0–2.2 $\times 10^7$ cells/ml for diploid strains) with media containing 15% glycerol or 7% DMSO. If lower than 50 OD_{600}/ml , centrifuge the pool at $500 \times g$ for 2 min. to pellet the cells, decant the supernatant, and resuspend to 50 OD_{600}/ml .
11. Aliquot 10–25 μl of pool into individually capped PCR tubes, and store at $-80^\circ C$. Pooled aliquots can be stored indefinitely.
12. If desired, perform a hybridization test and create a pool supplement to boost the signal for strains that do not give good signal when hybridized (*see Note 2*).

3.2. Pooled Growth

1. Before beginning, *see Subheading 4* for information about growth equipment options (*see Note 3*), using a pre-screen growth step to allow the strains to recover from freezing (*see Note 4*), choosing the number of generations of growth (*see Note 5*), selecting the correct dose for drug conditions (*see Note 6*), choosing a sufficiently large culture volume and starting OD (*see Note 7*), selecting the correct control condition (*see Note 8*), and replicating samples (*see Note 9*).
2. Thaw a frozen aliquot of pooled cells on ice.
3. Immediately dilute the pool into media containing conditions of interest, including a control condition. For drug experiments, if the drug added is dissolved in a solvent, an equal amount of solvent should be added to the control culture. Choose culture volumes, starting OD, and final OD using the guidelines in **Note 7**, or use one of the two options given below.
 - (a) *Option A*: Use a starting OD_{600} of 0.0625 in two 700- μl wells of a 48-well plate (*see Note 10*). Seal the plate with a plastic plate seal, rolling with a plate roller. If the condition requires aerobic growth (for example, non-fermentable carbon sources) poke a small hole in the membrane seal towards the side of each well to allow better aeration of the culture. Grow in a spectrophotometer at $30^\circ C$ with shaking set to the highest setting. Cells can be grown in a single culture for five generations.
 - (b) *Option B*: Use a starting OD_{600} of 0.0020 in a 50-ml culture in a 250-ml culture flask. Grow in a shaking incubator at $30^\circ C$ and 250 rpm. Cells can be grown in a single culture for ten generations.
4. Grow cells for the desired number of generations, or until the $OD_{600} = 2$. If cells have reached the desired number of generations, proceed to the next step. Otherwise transfer the cells to a fresh culture as in the previous step.

5. Collect 1–2 OD₆₀₀ of cells in a 1.5-ml tube. Spin the tubes at 500 g for 2 min in a microcentrifuge to pellet the cells, and remove the media.
6. If a starting cell sample (i.e., a “T0 time point”) is needed, add 1–2 OD₆₀₀ of pool from the freezer to a 1.5-ml tube. Spin the tube at max for 2 min in a microcentrifuge and remove the media.
7. If not proceeding to the next step immediately, store cell pellets at –20°C in a non-frost free freezer.

3.3. Purification and Amplification of Barcodes from Cell Samples

1. Purify genomic DNA from the cell pellets with the Zymo Research YeaStar kit using Protocol I (included with the kit), or another suitable method for purifying yeast genomic DNA. Elute the DNA with 300 µl 0.1× TE. Genomic DNA can be stored indefinitely at –20°C.
2. If desired, quantitate genomic DNA using a gel or a UV spectrophotometer. In practice, using a consistent volume of genomic DNA (we use 15 µl which contains ~0.1 µg) from the previous step without adjusting for variations in DNA concentration gives good results.
3. Set up two 60 µl PCR reactions for each sample, one for the uptags and one for the downtags (33 µl dH₂O, 6 µl 10× PCR buffer without MgCl₂, 3 µl 50 mM MgCl₂, 1.2 µl 10 mM dNTPs, 1.2 µl 50 µM Up or Down primer mix, 0.6 µl 5 U/µl Taq polymerase, ~0.1 µg genomic DNA in 15 µl).
4. Cycle as follows in a thermocycler with a heated lid: 94°C 3 min; repeat 30×: 94°C 30 s, 55°C 30 s, 72°C 30 s; 72°C 3 min, hold at 4°C. PCR products can be stored at –20°C.
5. If desired, check the resulting PCR products on a gel (*see Note 11*).
6. If not proceeding to the next step immediately, store PCR products at –20°C in a non-frost free freezer.

3.4. Array Hybridization and Scanning

1. These protocols use Affymetrix TAG4 arrays. *See Note 13* for more array platform options.
2. Set up a boiling water bath with a floating rack for 0.5-ml tubes and a slushy ice bucket. Set hybridization oven temperature to 42°C.
3. Fill the arrays with 90 µl 1× hybridization buffer.
4. Incubate at 42°C and 20 rpm for at least 10 min in the hybridization oven to pre-wet the array.
5. Prepare 60 µl hybridization mix per sample (45 µl 2× hybridization buffer, 0.3 µl B213 control oligonucleotide (1 fm/µl) (*see Note 14*), 7.2 µl of mixed oligonucleotides (12.5 pm/µl) (*see Note 15*), 1.8 µl 50× Denhardt’s Solution) immediately before using and aliquot into 0.5-ml tubes suitable for boiling.

6. While arrays are equilibrating, add 20 μl of uptag PCR and 20 μl of downtag PCR to 60 μl of hybridization mix for a total volume of 100 μl .
7. Clip a tube-holder onto the lid of each tube to prevent the tubes from opening during boiling.
8. Boil each tube for 2 min and then set on ice for 2 min.
9. Spin tubes briefly to bring the samples to the bottom of the tubes.
10. Remove hybridization buffer from the arrays.
11. Add the samples to the arrays (90 μl per array).
12. Place a Tough-Spot over each of the two gaskets to prevent evaporation.
13. Hybridize for 3–16 h at 42°C and 20 rpm. It is best to keep the hybridization time constant for samples that are part of the same dataset.
14. Prime the fluidics station: put the tubings in place, empty the waste bottle, fill wash A, wash B, and water bottles, and run fluidics station protocol PRIME_450.
15. Prepare 600 μl biotin labeling mix per sample (180 μl 20 \times SSPE, 12 μl 50 \times Denhardt's Solution, 6 μl 1% Tween 20 (vol/vol), 1 μl 1 mg/ml streptavidin–phycoerythrin, 401 μl dH₂O). Prepare this mix immediately before using.
16. Aliquot 600 μl biotin labeling mix per chip into tubes.
17. Remove Tough-Spots from chips.
18. Remove hybridization mix and fill chips with 90 μl wash A.
19. Wrap chips that are waiting to be washed in aluminum foil.
20. Under Experiments: for each chip type in the sample name and the experiment name, enter the barcode.
21. Under Fluidics: enter the station, and which chip you will wash and label in which module using the “experiment” pull-down menu. Use the protocol “Genflex_TAG4_wash_protocol.”
22. When the chips are ready for hybridization, check for air bubbles and wash again if necessary before you engage the wash block.
23. Clean the glass window on each array with isopropanol and a cotton swab or lint-free wipe.
24. Put Tough-Spots on the gaskets to prevent evaporation and put them in the scanner.
25. Scan at an emission wavelength of 560 nm.
26. When all fluidics operations are complete, put all cabling in Millipore water and run SHUTDOWN_450.

3.5. Analysis of Microarray Results

3.5.1. Outlier Masking

The Affymetrix TAG4 array contains at least five replicate features for each tag probe, dispersed across the array to make their performance as independent as possible (35). This allows outlier features to be identified and discarded before calculating an average intensity value for each tag (*see* **Notes 12 and 16**).

1. For each feature on the array, examine the features within the surrounding 5 feature \times 5 feature region. If at least 13 of the 25 probes in this region differ from their trimmed replicate mean (the mean of the three middle replicates, excluding the highest and lowest replicates) by more than 10%, consider this probe part of an outlier-dense region that is not suitable for data analysis.
2. Once these outlier-dense regions have been identified, pad them by also including any probes that are within a 5-probe radius, as defined by $((x_1 - x_2)^2 + (y_1 - y_2)^2)^{1/2} < 6$ where x_1 , x_2 , y_1 , and y_2 are the x and y coordinates for the two features.
3. Also discard features for which (standard deviation of feature pixels/mean feature pixels) > 0.3 . This unevenness is typically caused by bright pieces of debris on the chip that affect only a single probe, and are therefore not padded to include neighboring features. The standard deviation is included in the .cel file for Affymetrix arrays.
4. After identification and removal of outliers, intensity values are calculated for each tag by averaging all unmasked replicates.

3.5.2. Saturation Correction

Signal on the TAG4 array is not linearly related to tag concentration because of feature saturation (35). If left uncorrected, this saturation will cause the degree of sensitivity or resistance to be underestimated for strains with brighter tags. Saturation is corrected by comparing the uptag and downtag ratios (*see* **Note 17**).

1. Using a pair of arrays that are not sample replicates, calculate $\ln(i_c - \text{bg})/(i_t - \text{bg})$ for each tag, where i_c is the control intensity, i_t is the treatment intensity, and bg is the background as estimated by taking the mean intensity of the unassigned tag probes.
2. Mark the ratios for any tags with minimum values less than $3 \times$ background as unusable (this mark is for saturation correction only; a more robust threshold for unusable tags will be re-calculated later for use in the final results).
3. Pair uptag and downtag ratios by strain. Ignore ratios for any strains with less than two usable tags.
4. For each strain, calculate the difference in average intensity for the two tags: $(i_{\text{tu}} + i_{\text{cu}})/2 - (i_{\text{td}} + i_{\text{cd}})/2$, where the subscript u indicates the uptag, and the subscript d indicates the downtag.

5. Sort ratios by the difference in average intensity as calculated above. Take a sliding window of 600 ratio pairs, sliding 100 pairs at a time. For each window fit a line to the uptag ratios (x -axis) versus the downtag ratios (y -axis) using least-squares and take the slope. Also calculate the mean of the differences in the average intensity for the window.
6. Fit a least-squares line to the intensity differences for all windows (x -axis), versus slopes for all windows (y -axis) and take the slope of this line.
7. Repeat **steps 4–6** using the reverse intensity difference: $(i_{td} + i_{cd})/2 - (i_{tu} + i_{cu})/2$, and taking the slope with the axes reversed: the downtag ratios on the x -axis, and the uptag ratios on the y -axis.
8. Average the slope calculated in **step 6** with the slope calculated in the repeat of **step 6**. This is the saturation correction factor S . A typical range for S is 0.0001–0.0005.
9. Adjust the raw intensity data using the following transformation: $f(i) = ie^{is}$.
10. To correct more than two arrays, calculate S for all possible pairs of arrays, then use the median of these values as the correction factor for all arrays in the set. Using a larger group of arrays will improve the accuracy of S .

3.5.3. Array Normalization

The uptags and downtags should be normalized separately, because they are amplified in separate PCR reactions, and the intensities of the individual PCR reactions will affect the intensities of their array. Normalize half of the unassigned probes with the uptags, and half with the downtags. These probes are normalized and saturation corrected with the other probes to keep their values comparable. Uptag background and downtag background are calculated separately, as normalization may have a different impact on the background level for each tag set. Normalize array data using either quantile normalization (option A) (36) or mean normalization (option B) (see **Notes 18** and **19**).

1. *Option A*: To quantile normalize a set of arrays:
2. For each set of tags (up and down), rank values obtained from each array in the order of increasing intensity.
3. For each rank, assign the tag at that rank for each array to the median of all values at that rank.
4. *Option B*: To mean normalize a set of arrays:
5. For each set of tags (up and down), divide by the mean.
6. Multiply each tag set by the mean across all arrays (this is for convenience only; it returns the tag intensities to approximately their original range).

3.5.4. Removing Unusable Tags

Tags with low intensity values in the control samples will give poor-quality results. An intensity value threshold for excluding these tags should be chosen by comparing the correlation of uptag and downtag ratios as a function of tag intensity (*see Note 20*).

1. Using any treatment-control pair, calculate $\log_2(i_c - \text{bg}) / (i_t - \text{bg})$ for each tag, where i_c is the control intensity, i_t is the treatment intensity, and bg is the mean intensity of the unassigned tag probes for the appropriate tag set (uptag or downtag).
2. Pair uptag and downtag ratios by strain. Ignore ratios for any strains with only one tag (~200 strains only have an uptag).
3. For each pair, take the minimum intensity for the two tags in the two samples. Sort the ratio pairs by this minimum intensity.
4. Use a sliding window of size 50 on the ranked ratio pairs, starting with the lowest intensity pairs. Calculate the correlation of uptag and downtag ratios for pairs within the window. Also calculate the average of the minimum intensities calculated in the previous step.
5. Slide the window by 25 pairs, and repeat the previous step until all pairs have been traversed.
6. Plot the average minimum intensity versus the uptag–downtag correlation for all windows.
7. Chose an intensity threshold from this plot by eye, or by using a set formula. For example, the intensity value where the correlation first reaches 80% of its maximum level is a good cutoff.
8. Mark any tags that are below this cutoff in either of the samples as unusable for calculating log ratios. Mark any tags below this cutoff in any of the control samples as unusable for calculating p -values. The usability criteria are less strict for p -value calculation than for log₂ ratio calculation because log₂ ratios are a measure of how sensitive each strain is, whereas p -values are only a measure whether the strain is sensitive or not. It is possible for a strain that is above background in the control samples but below background in the treatment sample to be significantly sensitive, but without a usable treatment sample the degree of sensitivity cannot be measured.

3.5.5. Calculating p -Values and False Discovery Rates

This method works best for cases where a large set of common control arrays (more than ten) are used with one or two replicate arrays per experiment (*see Note 9*). A validation of this method based on a comparison to the number of false sensitivity calls in matched control–control comparisons has been described in detail (15). If multiple replicates are available for the control and the treatment samples and a large control set is not available, other methods for calculating significance such as those used in the Significance Analysis of Microarrays (SAM) package (37) are preferable (*see Notes 21–27* for more information about interpreting results).

1. For each usable tag, calculate the mean of the controls (μ_c) and the standard deviation of the controls (σ_c).
2. For each strain, average all usable tags. If a strain has no usable tags, this strain should be excluded from the analysis.
3. For each strain, calculate a z-score: $(\mu_c - t)/\sigma_c$ where t is the treatment intensity for that strain. Strains that are sensitive will have positive scores, while strains that are resistant will have negative scores.
4. Calculate p -values from the z-scores by fitting a t -distribution with $n_c - 1$ degrees of freedom to all scores for the experiment, where n_c is the number of control arrays.
5. Calculate a corresponding false discovery rate (FDR) for each of the p -values as follows. Rank the p -values in ascending order. For the p -value with rank r , the corresponding FDR is $(pns)/r$ where n_s is the total number of strains in the analysis.

3.5.6. Calculating log 2 Ratios

Whereas p -values describe the level of confidence for calling strain sensitivity, the log 2 ratios give the best estimate of the actual level of sensitivity for each strain. For example, deletions in the same complex or linear pathway will often cause similar growth defects, and as a result these strains will often have similar log 2 ratios. However, sensitivity for these strains may be measured with different degrees of accuracy, and they will therefore not necessarily have similar p -values (*see* **Notes 21–27** for more information about interpreting results, and **Note 28** for sample ratio results).

1. For each tag, calculate $\log_2(\mu_c - bg)/(\mu_t - bg)$ where μ_c is the mean intensity for the control samples, μ_t is the mean intensity for the treatment samples, and bg is the mean intensity of the unassigned probes for the appropriate tag set.
2. For each strain, average the log 2 ratios for all usable tags to obtain a final sensitivity score.
3. Strains that are sensitive will have positive scores, while strains that are resistant will have negative scores. This score is proportional to the log 2 ratio of cells present in the control sample versus the treatment sample.

4. Notes

1. *Strain background options:* The yeast deletion collection contains four different sets of strains: homozygous diploids, heterozygous diploids, $MATa$ haploids, and $MAT\alpha$ haploids. In general it is preferable to work with the heterozygous and the homozygous collections because the diploid genome is

more robust against phenotypic effects caused by secondary site mutations. These mutations are typically heterozygous in the diploid collections, whereas for the haploid collections a secondary site mutation will affect the only available gene copy.

2. *Adjusting strains with low tag intensities:* To identify strains that are underrepresented in the pool, perform a test hybridization. It is best to grow the pool in the standard control condition before this test hybridization, because slow growing strains often drop out of the pool during this growth step (15% of the homozygous deletion strains and 3% of the heterozygous strains are slow-growing when compared to wild type (1, 5)). Once identified, the signal for these strains can be boosted by adding extra cells. Strains that are not detectable can be added in larger quantities to a supplementary pool to boost their signal. It is normal that 200 strains or more remain undetectable even after interventions of this type.
3. *Growth equipment options:* There are many possible ways to grow pooled collections of deletion mutants. We grow our pooled cultures in shaking spectrophotometers that are connected to pipetting robots, allowing the cells to be inoculated into new cultures or harvested robotically when they reach the desired optical density (OD). This greatly facilitates the process and increases reproducibility by making the growth step as consistent as possible, however it is by no means necessary for obtaining good results. Cells can be grown in a shaking spectrophotometer to facilitate OD tracking and then transferred and harvested manually, or cells can be grown in flasks with both OD readings and transfers performed manually. Cells could also be grown as a lawn on solid media, or in alternate growth equipment such as a chemostat.
4. *Using a pre-screen growth step to allow the strains to recover from freezing:* Experimental cultures can either be inoculated directly from frozen aliquots or from a starting culture that has been grown for a set number of generations before starting the experiment. Including a recovery step allows the cells to recover from freezing, and may improve the accuracy of the sensitivity results. The main disadvantage to including this step is that increasing the total growth time causes more slow-growing strains to drop out of the pool by the final time-point, resulting in a greater number of strains with unusable data.
5. *Choosing the number of generations of growth:* In general, a longer growth period can more sensitively detect subtle phenotypes, but results in the depletion of slow-growing strains from the population (in both the experimental and control condition), thus precluding the analysis of these strains. Therefore, the ideal growth period for a pooled culture will

vary. For example, it is generally preferable to grow heterozygous strains longer than homozygous strains because heterozygous growth phenotypes are generally more subtle. For resolution across a wide array of sensitivities, cells can also be collected at multiple time-points.

6. *Selecting the correct dose for drug conditions:* In general, a stronger treatment dose will reveal more sensitive strains by making the inhibition of slightly sensitive strains easier to detect. However, a higher dose will also increase the number of strains that do not grow at all, precluding the ranking of these strains' sensitivities. In general, we aim for treatment doses that cause a 10–20% decrease in growth rate when tested on a wild type. Standardizing the level of inhibition allows most experiments to be sampled after the same number of generations of growth. At 10–20% inhibition, optimal results for the heterozygous collection are usually obtained at the 20 generation time point, and optimal results for the homozygous collection are usually obtained at the five generation time point. In general, a more benign treatment will cause smaller changes in growth rate, and will therefore require a longer growth period to resolve differences between strains.
7. *Choosing a sufficiently large culture volume and starting OD:* Choosing an appropriate culture volume is critical for obtaining good results. At each inoculation step, the cells are diluted to their lowest numbers. Decreasing the average number of cells per strain at inoculation increases the effect of variation on the actual number of cells sampled for each strain. This variation is typically the most significant source of noise in the assay (38). Using at least 300 cells per strain is recommended. The number of cells per strain at inoculation can be calculated as follows: Cells per strain = (Inoculation OD) × (culture volume (ml)) × (cells/ml at OD = 1)/(number of strains). A culture of diploid cells at OD₆₀₀ 1.0 has approximately 2.2×10^7 cells/ml. If the culture vessel will not accommodate the volume needed to reach the desired cell numbers, multiple cultures can be grown in parallel and pooled at the end of the experiment. Increasing the number of inoculations will also increase noise, and should be minimized. Also keep in mind the specific needs of the growth setup when choosing a culture volume (for example, manual OD readings will require extra culture volume).
8. *Selecting the correct control condition:* Raw values of array intensity for a single sample are very poorly correlated with the abundance in absolute strain. For this reason, each experiment requires at least two samples – a control sample and a treatment sample. The intensity ratio between two samples is a good measure of the abundance ratio of the corresponding strain (35,

38). The appropriate control condition will vary depending on the experiment. To measure relative growth rates, different time-points should be compared. To measure the relative sensitivity of each strain to a treatment condition, cells should be grown in a treatment condition and a control condition for the same number of generations. This type of comparison is useful because it takes each strain's normal growth-rate into account, preventing strains that are slow-growing in the absence of treatment, from appearing falsely sensitive.

9. *Sample replicates*: Each experiment requires at least two arrays – one for a control sample, and one for a treatment sample. A single treatment-control pair gives good-quality data, especially for cases where the data set taken as a whole is more important than the accuracy of any single experiment. For example, in cases where the results of many experiments are used to build a gene network, the results will be relatively robust against false positives, and testing more conditions may improve the results more than increasing the number of experimental replicates. For cases where individual gene sensitivity predictions are important, replication may also be unnecessary if array experiments are confirmed with single-strain follow up tests. For more accurate data, two or more replicates of both the control and treatment conditions are recommended to allow for the identification and removal of rare outlier points. Increasing the number of replicates will also allow for more subtle phenotypes to be detected with statistical significance. We often use a large, shared set of control arrays (more than ten) for analyzing many different experimental arrays, each with only one replicate. This control set can be used to calculate the statistical significance of the final results, while minimizing the total number of experimental arrays needed. If using a large, shared control set it is important that the control arrays represent as diverse a set of samples as the treatment arrays (cells grown on different days, tags amplified in different runs, etc.) so that the variation in the control set will accurately represent the non-treatment related variation present in the treatment arrays. When not using a shared set of control arrays, it is best for each control-treatment pair to be processed together (cells grown on the same day, tags amplified in the same run, etc.) to minimize the non-treatment related differences between the two. As discussed in **Note 7**, the growth step is the main source of noise in the assay. For this reason, all replicate samples should be full biological replicates (grown separately, as opposed to separate hybridizations of the same growth sample). It is important to note that the uptag and downtag are not independent measures of strain sensitivity. The two tags will share the noise generated during the growth step, where any sampling error affects both tags simultaneously.

The uptag and downtag therefore can not be treated as independent sample replicates when using statistical analysis methods. All useable tags for each strain should therefore be averaged before using these methods, with only true experimental replicates treated as replicate samples.

10. *Comparing OD₆₀₀ values between plates and cuvettes:* If growing cells in a shaking spectrophotometer, note that the OD₆₀₀ measured for the plate will differ from the OD₆₀₀ of the same culture measured in a cuvette due to differences in path length. Similarly, OD₆₀₀ readings in a shaking spectrophotometer will also vary with differences in culture volume. All OD₆₀₀ values given in this paper refer to those measured in a 1 ml, 1-cm path length cuvette.
11. *Checking for successful tag amplification:* To check for successful amplification, tagged PCR products can be separated on a gel. The desired product is ~60 bp. A second band is often seen because non-complementary tags can hybridize at their common primer regions, forming a partially single-stranded structure that migrates more slowly than the fully double-stranded tag products. Reactions that appear unusually dim (less than 5× the concentration of a typical reaction) will often give low-quality array data. These amplifications can be repeated using the leftover genomic DNA. It is common to have amplification in no-template control reactions because tag contamination is extremely difficult to prevent. This low-level of contamination will typically not impact results because the concentration of the experimental template will be much greater than the concentration of contamination in the experimental samples, preventing the contamination from having a serious impact on the results.
12. *Barcode sequences:* The deletion of a given gene will use the same two barcodes in all four strain backgrounds. Because of this barcode overlap, the homozygous deletion collection and the heterozygous deletion collection must be pooled and grown separately to ensure that each barcode corresponds to a unique strain within the pool. A full list of sequences is available (38). It is an Affymetrix convention to list probe sequences for their arrays in 3'→5' orientation because this is the direction of probe synthesis used in their manufacturing process. For this reason, the tag sequences listed in Affymetrix array files are the reverse of those listed elsewhere. This is a common source of confusion, so sequence orientation should be carefully checked when designing new barcoded strains or other samples for use with Affymetrix tag arrays.
13. *Array options:* These protocols use the Affymetrix tag arrays; however they can easily be adapted to the array platform of your choice. For alternative array examples see refs. 4, 12, 13, 39, 40.

14. *Determining the correct B213 concentration:* It is common when diluting a new batch of B213 control oligo to have to test several concentrations to find a concentration that gives good results. The reason for this inconsistent performance from batch to batch is unknown. The B213 oligo binds to the border probes on the array, which are used by the Affymetrix software to align the grid that defines the probe borders. If the control spots are too dim, automated alignment will fail, and if the control probes are too bright, they can interfere with neighboring probes on the array. It is best to aim for an intensity that is bright enough for alignment to work consistently, but within the range of the tag probe intensities. If grid alignment fails because the border probes are too dim, it is possible to align the grid manually and recalculate the probe intensities using the Affymetrix software.
15. *Mixed oligos:* The mixed oligos are added to prevent re-annealing of the tagged product after melting. Non-complementary tags can stick together by their common primer sequences, forming a partially single-stranded product. Oligos complementary to the common primers are added in excess to prevent the formation of this product, making the tag sequences more accessible for hybridization.
16. *Re-hybridizing failed samples:* The PCR volume recommended here provides enough PCR for two hybridizations. If the first hybridization has problems such as an extremely uneven appearance (for example, due to large bubbles during washing or staining) or an unusually large amount of visible debris or damage, the remaining PCR can be hybridized to a new array. The replicate features can correct for minor damage, so in general, only damage that affects more than approximately one fifth of the array surface is a concern.
17. *Saturation correction:* The saturation correction function is derived from the data by exploiting the fact that the raw intensities for a strain's uptag andowntag often differ, despite the fact that each uptag–owntag pair is measuring the same true strain ratio. The saturation pattern can therefore be derived by comparing the difference in raw intensity to the difference in ratio for all uptag–owntag pairs. The data is then corrected to make the ratios independent of raw intensity. This transformation increases the correlation of uptag andowntag ratios. The correction algorithm requires at least two arrays, and works best with a pair of arrays that are not sample replicates. This is because all log ratios for a pair of sample replicates will be approximately zero, and saturation affects ratios close to zero the least. The level of saturation can vary from day to day because of differences in hybridization time, washing, and staining time. For this

reason it is best to derive the correction function for all arrays hybridized on a given day as a group, and avoid using pairs of arrays processed on different days for deriving the correction function.

18. *Normalizing array data:* Two alternate normalization methods are discussed here – mean normalization, and quantile normalization (36). Quantile normalization is often preferred because it can normalize data effectively even when there is a non-linear relationship between the samples, but this method requires the assumption that the overall distribution of tag intensities is the same for all experiments being normalized. Mean normalization, though only capable of linear transformation of the data, does not require this assumption. In general, quantile normalization works best for experiments in which a reasonably small number of strains are affected by the condition tested. For experiments with a large number of sensitive strains, mean normalization often performs better. In particular, experiments with different growth periods (for example, the zero time-point and the 20 generation time-point) or from different starting pools (for example, homozygous pools constructed separately) should not be normalized together using quantile normalization because the tag intensities for these samples will have very different distributions (38).
19. *Normalizing new arrays without renormalizing old data:* A slight variation on the normalization procedures described can help make data analysis more convenient by allowing new data to be normalized without having to re-normalize all existing data. When quantile normalizing a group of arrays, a standard curve is usually derived from the all arrays in the group. If you wish to keep the standard curve constant over time, it can be calculated from a starting set of arrays and then held constant for normalizing future arrays. This works well as long as all arrays normalized are of the same experimental type (same pooled cells, grown for the same number of generations). Similarly, when mean normalizing a set of arrays, the mean intensity across all arrays in the set is used to return the mean-normalized values to their original range. This mean can also be calculated from an initial set of arrays and then held constant.
20. *Removing unusable tags:* When tags become too dim to be usable, their correlation with their tag replicates degrades to zero. By comparing the correlation of uptag and downtag ratios to tag intensity, an effective cutoff can be chosen to remove these tags. In practice, this intensity cutoff remains relatively constant across experiments as long as the experimental setup (in particular the hybridization time and the scanner used) does not change, so once a threshold is chosen, it is not required to be recalculated with every experiment.

21. *Reasons expected hits may be absent:* If a strain for which a sensitive phenotype was expected is not called as significantly sensitive, a good first thing to check is what the maximum achievable score is for that strain based on the control experiments. For example, if a strain is already fairly close to the background in the controls, or if the control data for that strain is noisy, it may not be possible for the experimental sample to be dim enough to give a significant score. To calculate the maximum achievable score for each strain, create a mock experimental chip in which all of the tag values are set to background, which is the minimum possible experimental value for each strain. Calculate scores using this mock-experimental chip and the control set, and apply the same FDR cutoff used for the true experimental results. This will give the maximum score achievable for each strain. If a strain is not called as sensitive in this mock experiment, then the data for that strain can be considered inconclusive. It is also possible that the pooled growth phenotype for the strain of interest differs from its phenotype in another environment. Growth in a shaking spectrophotometer, growth in a flask, and growth on a plate may be slightly different because of differences in these growth environments. For example, a culture in a shaking spectrophotometer typically has a lower oxygen level than a standard flask culture. Pooled growth may also create differences in phenotype. A deletion strain grown in a pooled culture is predominantly surrounded by strains that are wild type at its deleted locus. Under certain conditions these surrounding strains may be able to compensate for the deficiencies caused by the deletion.
22. *Evaluating data quality and replicating sample agreement:* The most effective way to measure the quality of replicate samples is to measure the correlation of the log-transformed, normalized, and saturation-corrected intensity values, termed “processed raw values.” This correlation should always be at least 0.98 for replicate samples. Processed raw values are a better indicator of sample quality than raw values because normalization and saturation correction remove artifacts that will artificially decrease replicate correlation, and log transformation adjusts the values so that a given deviation from the diagonal will have a consistent impact on the final log ratios, regardless of tag intensity. Lower correlation may indicate a sampling error problem, which can be corrected with increased culture volume as described in **Note 7**. Uptag andowntag ratios for a given sample will be positively correlated; however it is important to note that low correlation for uptag andowntag ratios does not necessarily indicate poor sample quality. In experimental samples where few strains are sensitive, most of the ratios are centered at zero, causing low correlation. In perfect data, the uptag–owntag log₂ ratio correlation for a control–control pair would be

zero, but sampling error causes slight strain representation differences between replicates, and the two tags correlate in measuring this noise. Replicate samples for which few strains are sensitive will also have low log 2 ratio correlation for the same reason. Because of this property, log 2 ratio correlation is a poor measure of replicate sample agreement.

23. *Comparing results across different treatment conditions:* The same score in two different experiments does not indicate the same level of inhibition relative to the control culture. To understand why, note that for any given condition, the phenotype for the majority of gene-deletion mutants is not significantly different from that of the wild-type control strain, and the median score for the experiment can therefore be assumed to be approximately that of the wild-type strain. This median score is also typically close to zero, because the final sample always contains the same number of cells no matter how slow growing a culture is. Therefore, for an experiment where the treatment culture grew 20% slower than the control culture, strains with a log 2 ratio of zero also grew approximately 20% slower than in the control culture. Similarly, in a case where the treatment culture grew 15% slower than the control culture, strains with a log 2 ratio of zero grew approximately 15% slower than in the control culture. Therefore, when comparing treatment inhibition for the same strain across multiple treatments, the growth rate of each pooled culture needs to be taken into account. Theoretically it should be possible to adjust log 2 ratios by multiplying the scores for each experiment by the average level of inhibition for each culture, however this idea has not been tested in practice.
24. *Reduced data quality for slow growing strains:* The data for strains with lower-than-average representation in the pool is noisier due to increased sampling error. Strains that are slow-growing in the control condition are especially prone to this problem because they often have low representation in the pool (5). Data from these strains should be carefully confirmed. It is also common for slow-growing strains to appear slightly resistant to treatments that cause a large decrease in growth rate. While the reason for this effect is unknown, it is thought that slow growth may be less limiting in a condition where the pool as a whole is growing more slowly, causing these strains to be slowed proportionately less than others and thereby appear resistant.
25. *Frequently sensitive strains:* Some gene deletions are known to cause sensitivity in a large number treatment conditions (13, 15). While the sensitivity of these deletions is genuine, strains specifically sensitive to the condition of interest may

be better candidates for follow-up experiments than genes known to mediate sensitivity to a broad range of conditions. Common gene families include: ergosterol biosynthesis, aromatic amino-acid biosynthesis, genes related to transcription, and genes with a variety of membrane functions such as ER to Golgi transport, vesicle mediated transport, and vacuolar targeting (5, 15). The frequency of sensitivity for each strain in a large number of published experiments can be used as a reference when evaluating results (15).

26. *Interference from neighboring gene deletions:* Gene function can be disrupted by deletion of a neighboring gene, either by partial deletion of the ORF due to overlap, disruption of promoter regions, or by generation of secondary site mutations during homologous recombination. It is common to see pairs of adjacent genes that are both sensitive to the same conditions. The true gene responsible for the phenotype can be determined by adding each of the two genes back to each of the two deletion strains on a low copy-number plasmid. The gene that is responsible for the phenotype should complement the deletion in both of these strains.
27. *Interpreting data for drug-target identification:* In addition to the deletion strain corresponding to the drug target, there may be other heterozygous deletions that are sensitive. For example, a gene involved in coping with the stress caused by the drug may also be sensitive due to reduced protein level. Because of this, there is no set way for determining which, if any, of the sensitive strains corresponds to the drug target from the heterozygous profiling data alone. Other data types are often helpful for narrowing the candidates. Synthetic lethality data can aid in interpreting the results (12, 13, 39, 41, 42). Homozygous deletions corresponding to synthetic lethal partners of the drug target should be more sensitive to the drug because the drug mimics a deletion (or partial deletion) of the drug target. Synthetic lethality data is available for many complete deletions and also for essential genes in the form of promoter replacement alleles (43). and 3' untranslated region disruption alleles (44). both of which reduce gene dosage. Both of these data types are useful for clarifying heterozygous profiling results. Genetic interaction data can be found in the BioGRID database (45). Genome-wide overexpression data is also useful, as drug resistance often results from overexpression of the gene target (17). In addition, the corresponding homozygous deletion data (if a full deletion of the gene of interest is viable) should be examined. If both the heterozygote and the homozygote are sensitive the gene is unlikely to encode the drug target because a strain that lacks the drug target should not be affected by the drug.

28. *Sample results:* Fig. 6 shows sample data for an experiment in which the amino acid lysine was omitted. This data shows the typical level of agreement for raw data from multiple experiments, and the typical transformation of the raw data caused

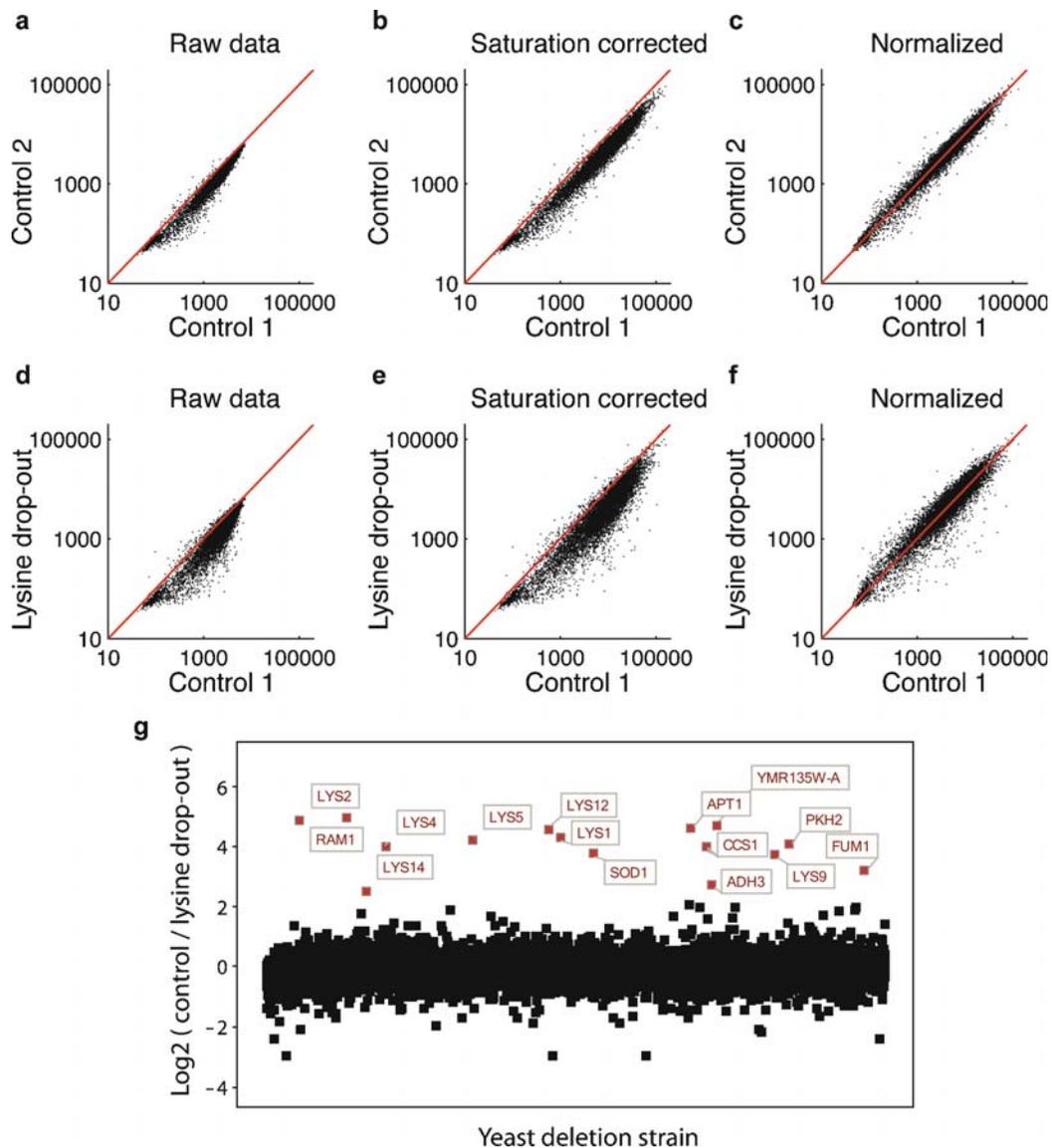


Fig. 6. Sample data showing data analysis steps and expected results. Sample data is shown for a control–control pair, and for a lysine drop-out treatment sample compared to a control. (a) Raw tag intensities for the control–control pair (replicate array features have been averaged for each tag). (b) Data after saturation correction. Note that the range of intensities increases. (c) Data after quantile normalization. Note that the data is now centered along the diagonal. (d–f) The same steps are shown for the lysine drop-out treatment versus a control. (g) \log_2 intensity ratios for the lysine drop out experiment. Positive values indicate sensitivity in the treatment condition, negative values indicate resistance. As expected, genes known to be involved in lysine biosynthesis are overrepresented among the sensitive strains ($p = 1.13 \times 10^{-14}$). Modified from ref. 38.

by saturation correction and quantile normalization. Data is shown for both a pair of controls and a control-treatment pair. As expected, the raw data for the pair of control samples is more similar than the raw data for the control-treatment pair. Final log₂ ratios are shown for the lysine drop-out condition versus the control condition. As expected, the majority of sensitive strains have known roles in lysine biosynthesis. Note that the log₂ ratios are similar for these strains, suggesting that they have a similar level of sensitivity.

Acknowledgments

Work In the Giaever and Nislow labs is supported by the NHGRI, the Canadian Foundation for Innovation, and the CIHR (MOP-81340 to GG) and (MOP-84305 to CN).

References

- Giaever, G. et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–91 (2002).
- Winzler, E.A. et al. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–6 (1999).
- Shoemaker, D.D., Lashkari, D.A., Morris, D., Mittmann, M. & Davis, R.W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat Genet* **14**, 450–6 (1996).
- Birrell, G.W. et al. Transcriptional response of *Saccharomyces cerevisiae* to DNA-damaging agents does not identify the genes that protect against these agents. *Proc Natl Acad Sci USA* **99**, 8778–83 (2002).
- Deutschbauer, A.M. et al. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**, 1915–25 (2005).
- Giaever, G. et al. Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc Natl Acad Sci USA* **101**, 793–8 (2004).
- Giaever, G. et al. Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat Genet* **21**, 278–83 (1999).
- Kastenmayer, J.P. et al. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res* **16**, 365–73 (2006).
- Lee, W. et al. Genome-wide requirements for resistance to functionally distinct DNA-damaging agents. *PLoS Genet* **1**, e24 (2005).
- Lum, P.Y. et al. Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell* **116**, 121–37 (2004).
- Ooi, S.L., Shoemaker, D.D. & Boeke, J.D. A DNA microarray-based genetic screen for non-homologous end-joining mutants in *Saccharomyces cerevisiae*. *Science* **294**, 2552–6 (2001).
- Parsons, A.B. et al. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat Biotechnol* **22**, 62–9 (2004).
- Parsons, A.B. et al. Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* **126**, 611–25 (2006).
- Steinmetz, L.M. et al. Systematic screen for human disease genes in yeast. *Nat Genet* **31**, 400–4 (2002).
- Hillenmeyer, M.E. et al. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–5 (2008).
- Hoon, S. et al. An integrated platform of genome-wide assays reveals small molecule bioactivities. *Nat Chem Biol* **4**(8), 498–506 (2008).
- Ericson, E. et al. Off-target effects of psychoactive drugs revealed by genome-wide assays in yeast. *PLoS Genet* **4**(8), e1000151 (2008).

18. Workman, C.T. et al. A systems approach to mapping DNA damage response pathways. *Science* **312**, 1054–9 (2006).
19. Jensen, L.J., Jensen, T.S., de Lichtenberg, U., Brunak, S. & Bork, P. Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature* **443**, 594–7 (2006).
20. Pollack, J.R. et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA* **99**, 12963–8 (2002).
21. Groh, J.L., Luo, Q., Ballard, J.D. & Krumholz, L.R. A method adapting microarray technology for signature-tagged mutagenesis of *Desulfovibrio desulfuricans* G20 and *Shewanella oneidensis* MR-1 in anaerobic sediment survival experiments. *Appl Environ Microbiol* **71**, 7064–74 (2005).
22. Karlyshev, A.V. et al. Application of high-density array-based signature-tagged mutagenesis to discover novel *Yersinia* virulence-associated genes. *Infect Immun* **69**, 7810–9 (2001).
23. Berns, K. et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–7 (2004).
24. Brummelkamp, T.R. et al. An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors. *Nat Chem Biol* **2**, 202–6 (2006).
25. Fischer, K.D. et al. Defective T-cell receptor signalling and positive selection of Vav-deficient CD4 + CD8 + thymocytes. *Nature* **374**, 474–7 (1995).
26. Fraser, A. RNA interference: human genes hit the big screen. *Nature* **428**, 375–8 (2004).
27. Kolfschoten, I.G. et al. A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. *Cell* **121**, 849–58 (2005).
28. Ngo, V.N. et al. A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106–10 (2006).
29. Ngo, V.N. et al. A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106–10 (2006).
30. Westbrook, T.F. et al. A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121**, 837–48 (2005).
31. Akhras, M.S. et al. PathogenMip assay: a multiplex pathogen detection assay. *PLoS ONE* **2**, e223 (2007).
32. Clayton, D.G. et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* **37**, 1243–6 (2005).
33. Hardenbol, P. et al. Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat Biotechnol* **21**, 673–8 (2003).
34. Hardenbol, P. et al. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res* **15**, 269–75 (2005).
35. Pierce, S.E. et al. A unique and universal molecular barcode array. *Nat Methods* **3**, 601–3 (2006).
36. Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–93 (2003).
37. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98**, 5116–21 (2001).
38. Pierce, S.E., Davis, R.W., Nislow, C., & Giaever, G. Genome-wide analysis of bar-coded *S. cerevisiae* gene-deletion mutants in pooled cultures. *Nat Protoc* **2**, 2958–74 (2007).
39. Pan, X. et al. A robust toolkit for functional profiling of the yeast genome. *Mol Cell* **16**, 487–96 (2004).
40. Yuan, D.S. et al. Improved microarray methods for profiling the Yeast Knockout strain collection. *Nucleic Acids Res* **33**, e103 (2005).
41. Tong, A.H. et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–8 (2001).
42. Tong, A.H. et al. Global mapping of the yeast genetic interaction network. *Science* **303**, 808–13 (2004).
43. Davierwala, A.P. et al. The synthetic genetic interaction spectrum of essential genes. *Nat Genet* **37**, 1147–52 (2005).
44. Schuldiner, M. et al. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–19 (2005).
45. Stark, C. et al. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535–9 (2006).