Published in final edited form as:

Methods Mol Biol. 2015 ; 1263: 299–318. doi:10.1007/978-1-4939-2269-7_23.

Chemical genomic profiling via barcode sequencing to predict compound mode of action

Jeff S. Piotrowski¹, Scott W. Simpkins², Sheena C. Li³, Raamesh Deshpande², Sean McIlwain¹, Irene Ong¹, Chad L. Myers², Charlie Boone⁴, and Raymond J. Andersen⁵

¹ University of Wisconsin-Madison, Great Lakes Bioenergy Research Institute, Madison, Wisconsin, USA

² University of Minnesota-Twin Cities, Department of Computer Science and Engineering, Minneapolis, Minnesota, USA

³ RIKEN Center for Sustainable Resource Science, Wako, Saitama, Japan

⁴ University of Toronto, Terrence Donnelly Centre for Cellular and Biomolecular Research, Toronto, Ontario, Canada

⁵ Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada

Summary

Chemical genomics is an unbiased, whole-cell approach to characterizing novel compounds to determine mode of action and cellular target. Our version of this technique is built upon barcoded deletion mutants of *Saccharomyces cerevisiae* and has been adapted to a high-throughput methodology using next-generation sequencing. Here we describe the steps to generate a chemical genomic profile from a compound of interest, and how to use this information to predict molecular mechanism and targets of bioactive compounds.

Keywords

Chemical genomics; barcode sequencing; functional genomics; yeast deletion collection

1. Introduction

Chemical genomics is a powerful technique for understanding the mode of action and cellular targets for unknown compounds (1–4). The particular strengths of chemical genomics are that it is a whole cell assay that is not designed around a target of interest, and it provides an unbiased view of the cellular response to a compound (3). The technique is based on exposing a large pool of defined gene deletion mutants to a compound and

Correspondence: jpiotrowski@wisc.edu.

⁸The pools in glycerol can be aliquoted and stored at –80 °C. As very few cells will be used per experiment (especially if they need dilution), it's best to aliquot in small volumes (200–500 µL) so to extend the life of the stock collection.

⁹Do not exceed 1% solvent in the cultures. If resources allow, plan on using multiple dose points so that a “dose dependent” chemical genomic profile can be generated. Inhibition of growth by <20% of solvent controls may still yield information given the performance of individual mutants in the pool.

measuring the fitness of the individual mutants (5). The fitness of these mutants can be measured in a number of ways (*e.g.* colony size, optical density, etc). Several mutant collections have been created where the gene mutation is replaced with a specific molecular barcode, a short section of DNA with a sequence specific to the mutant (6–8). In the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, these barcodes are 20 bp sequences flanked by common priming sites that allow amplification of the barcodes. Because each barcode is unique, the mutants can be pooled together, and the relative fitness of each strain can be determined by the abundance of the mutant specific barcode, either by microarray or next-generation sequencing of the barcodes (9). Given that most genes in *S. cerevisiae* have been functionally annotated, the fitness of the mutants in the presence of a compound can give functional insight into the chemical's mode of action. For instance, in the presence of a DNA damaging agent like methyl methanesulfonate (MMS), mutants in genes involved in DNA damage and repair (*e.g. rad51*) are significantly more sensitive to the compound compared to other mutants (1, 5). The pool of mutants can also yield valuable knowledge for weakly bioactive compounds, as mutant performance may be dramatic even when the behavior of wild type (WT) cells appears unaffected (4). This ‘chemical genomic profile’ of the mutant collection, can further be paired with the yeast genetic interaction network to predict the pathways and proteins that the compounds may be directly affecting (10).

Chemical genomic profiling uses a pooled yeast deletion collection that is created by pipetting individual, liquid mutant cultures into a common pool. The pool of yeast mutants is central to all assays, thus careful construction of the pool is essential. Each time a pool is created, the starting distribution of mutants can generate a “pool signature” created by the distributions of mutants in the pool; therefore, it is best to make a large pool to cover more than the number of planned screens so the mutant pool does not have to be remade (*see Note 1*). Chemical genomics can be performed using individual mutants arrayed on agar and measuring colony size to determine fitness, but the compound requirements for such assays are orders of magnitude greater than pooled assays. As novel compounds are often scarcely available, we have focused on optimizing the liquid assay to minimize compound requirements.

The major strength of chemical genomics using barcode sequencing is the ability to multiplex samples and screen many different compound conditions in a single sequencing reaction. Present sequencing technology offers a high number of reads, allowing multiple samples to be multiplexed in a single sequencing lane and later de-multiplexed via sample-specific index tags built into the PCR primers. The maximum degree of multiplexing is determined by the sequencing platform and the size of the mutant pool. When using the entire non-essential, haploid yeast knock out collection (~5000 strains), a maximum of approximately 25 samples can be pooled and sequenced simultaneously on the Illumina MiSeq platform (using present flowcell design “V2”), while a up to 96 or more samples can be sequenced on a single HiSeq 2500 lane and 192 samples in a HiSeq 2500 Rapid Run (2

¹If purchased, the yeast deletion collection may arrive as glycerol stocks. These will need to be thawed and pinned to YPD+G418 agar before starting the pool creation. Strain pools can be made either by scraping colonies off agar (11) or mixing liquid cultures. We have tested both methods and have found that liquid cultures give a more equal distribution of strains (**Figure 3**).

lanes). A single Illumina, 8-lane can accommodate nearly 1000 chemical genomic screens and potentially much more as sequencing technology rapidly improves. Experiments should be designed with the multiplexing limitations of the sequencing platform in mind.

In this chapter we will describe the steps (**Figure 1**) necessary to perform chemical genomic analysis using the yeast non-essential gene deletion collection. We will describe creating the pool of yeast mutants and exposing these to a compound as a pooled competition. We then describe the steps to remove and amplify the molecular barcodes using multiplexed primers and then sequence these using next-generation sequencing. Finally we provide the computational tools and describe the steps to generate a chemical genomic profile and identify sensitive and resistant mutants in the presence of a chemical compound, which yields functional insight into the compound's mode of action.

2. Materials

When working with live cells, always follow proper sterile technique and use a laminar flow hood for all cell transfers. Media and reagents can be stored at 4 °C unless otherwise indicated.

2.1 Arraying the deletion collection to agar

1. YPD+G418 agar: Make 2 L. For each liter of the medium, add 10 g of yeast extract, 20 g of peptone, 20 g of agar and 950 mL of dH₂O and autoclave. To the cooled medium add 50 mL of 40% glucose (40 g glucose dissolved in 100 mL of H₂O and autoclaved) to make 1 L of 2% glucose YPD. Once cooled to 50 °C, add 1 mL of 1000× G418 (Geneticin, 200 mg/mL in dH₂O, filter sterilized) per liter.
2. 60 YPD+G418 square agar plates: Fill the plates with molten YPD+G418 agar in a sterile hood using a sterile pipette tool. Fill each plate with 35 mL of agar and let solidify.
3. Yeast non-essential deletion collection: If purchased, the collection will arrive in frozen glycerol stocks and need to be arrayed onto agar. The arrayed collection can be store at 4 °C for 1-3 months.
4. Pipetting tool and 50 mL sterile pipettes.
5. 96-well pin transfer tool or multichannel pipette.

2.2 Creating the pooled deletion collection for screening

1. YPD+G418 liquid medium: Make 2 liters. For each liter, add 10 g yeast extract, 20 g peptone and 950 mL dH₂O and autoclave. To the cooled medium add 50 mL of 40% glucose (40 g glucose dissolved in 100 mL H₂O and autoclaved) to make 1 L of 2% glucose YPD. Once cool to 50 °C, add 1 mL of 1000× G418 (Geneticin, 200 mg/mL in dH₂O, filter sterilized) per liter. This medium will be used to grow the mutants prior to creating the screening pool. Medium can be stored for up to 1 month at 4 °C.
2. 96-well, flat-bottom plates.

3. 96-well plate shaker.
4. Sterile 100 mL reservoirs.
6. 25-50 mL pipettes.
7. 2 L sterile flask with cover.
8. Centrifuge with a capacity for vessels with a volume >50 mL.
9. Spectrophotometer.
10. Microscope and haemocytometer.
11. 30% glycerol: Add 300 mL of glycerol to 700 mL dH₂O, autoclave.
12. Freezer tubes (1-2 mL) and boxes.

2.3 Components for pooled competitions

1. Screening medium: For basic screening use YPD with 2% glucose. For 1 L, suspend 10 g of yeast extract, 20 g of peptone in 950 mL of dH₂O and autoclave. Fill to 1 L with a sterile 50% glucose solution (*see Note 2*).
2. Compound(s) of interest in solution, control compounds (*e.g.* Benomyl, MMS), and solvent for control conditions. (*see Note 3*)

2.4 Components for genomic extraction, PCR, gel extraction, and template quantification

1. Genomic DNA extraction kits (*see Note 4*): For 1-96 samples use individual genomic extraction kits. For >96 samples, 96-well genomic extraction kits are preferred.
2. Zymolyase solution (only for 96-well genomic extractions): 1 mg/mL 100T zymolyase in 1 M sterile sorbitol (182.17 g sorbitol in 1 L, autoclaved).
3. Taq Super Mix: Contains Taq, dNTPs, buffer.
4. TE Buffer: 10 mM Tris-HCl, 1 mM EDTA, bring to pH 8.0 with HCl.
5. Indexed primer collection and common primer (*see Note 5, Table 1*): Prepare Indexed PCR primers, to a final concentration of 12.5 μ M and 100 μ M for the common U2 primer using TE buffer.
6. Agarose gel (2%).

²The assay described here is optimized for microcultures, which can be used in small-scale or high-throughput systems. Any growth medium can be used for the chemical genomics assays. We generally use rich media (*e.g.* YPD or YP Galactose with 2% sugar). We have found that using YP-galactose slows cell growth and slightly sensitizes the yeast to compounds, which can yield good separation of compound-responsive mutants in the pooled competition (14).

³Aside from water, DMSO is the preferred solvent as it has low toxicity to yeast. Stock solutions of the compound can be prepared depending on compound availability. A stock solution of 1 mg/mL is a good starting point to assess bioactivity of the compound.

⁴Depending on the scale, genomic DNA extraction can be performed individually or 96 wells at a time. Successful DNA extraction with the 96-well kits requires a pre-incubation of the cell pellet with zymolyase, as these kits are usually not specifically designed for yeast genomic extractions. Resuspend the cell pellet in 125 μ L of zymolyase solution and incubate for 1 h at 37 °C, after which the cells are ready for genomic extraction following the kit specifications.

⁵Our primers are designed and optimized for the Illumina platform. We have built in 10 bp Index Tags so that the experiments can be multiplexed. We have tested a range of index tag sizes, and found 10 bp to perform well.

7. Agarose gel extraction kit.
8. Kapa Illumina qPCR kit.

2.5 Sequence analysis and target prediction

1. Barseq counter software script package (available at www.github.com/csbio/barseq_counter)
2. Computer with Python (Version 2.7 or higher)
3. Computer with R-package (Bioconductor), EdgeR package, limma package, and corrplot package

3. Methods

3.1 Arraying the deletion collection to agar

Before the deletion collection can be pooled, the individual mutants must be grown in liquid culture. If purchased, the collection will arrive as frozen glycerol stocks. Rather than inoculate the liquid media directly from the glycerol, it is best to array the collection onto agar from the glycerol stocks first. This has the advantage of giving a ‘working collection’ that can be used for single mutant validations, or make new pooled collection without having to repeatedly thaw the deletion collection stocks and affect viability. The ‘working collection’ can be stored at 4 °C for 1-3 months before it should be transferred. This section describes making an agar array of the deletion collection from the glycerol stocks.

1. Thaw the frozen deletion collection glycerol stocks completely.
2. In a sterile hood, use the 96-well transfer device to spot a small volume (1-2 µL) of the glycerol stocks onto the agar plates. If using a non-disposable transfer device, be sure to bleach for 5 min and flame 3× between transfers to prevent contamination. Always be mindful of plate orientation (*i.e.* location of the A1 well).
3. Let the plates dry completely and incubate for 48 h at 30 °C.

3.2 Growing the yeast deletion collection to create the screening pool

To create the screening pool, each member of the deletion collection is cultured in independent wells and then pooled by pipetting the cultures together into a common pool. This section describes how to culture members of the deletion collection prior to pooling.

1. Working in a sterile hood, pipette 200 µL of YPD+G418 medium into each well of 60 sterile, 96-well flat bottomed plates. The purchased deletion collection in 96-well format is arrayed onto 57 plates. It is best to make 60 plates in case of accidents.
2. Using a 96-well transfer tool, transfer a small amount (about the size of a pin head) of cells from each agar array to a corresponding liquid media plate (*see Note 6*). If

⁶Make sure to maintain proper orientation of all plates. It is useful to mark the top left corner (the A1 well) of each plate before you begin.

not using disposable transfer tools, be sure to bleach the tips for 5 min and flame 3× between transfers.

3. Shake the liquid plates gently on a plate shaker for 30 s to distribute the cells.
4. Incubate the liquid cultures for 48 h at 30 °C until the cultures reach saturation (OD ~1.5)

3.3 Pooling the deletion collection

After the cultures have grown they are ready to mix together to make the screening pool for chemical genomic profiling. Making the screening collection is a sensitive step, and care must be taken to ensure the screening pool is at a high cell density with an even representation of strains. This step describes how to pool the individual cultures to ensure a homogeneous mixture, and importantly how to adjust the cell density to have adequate strain representation in downstream pooled competition assays.

1. In a sterile hood and using a multichannel pipette, remove the entire volume of the saturated liquid cultures from the 96-well plates and expel them into a reservoir. Pipette up and down or mix the plates on a plate shaker to ensure the cells are all in suspension before removal.
2. When the reservoir is full, use a 25 (or 50) mL pipette to transfer the liquid culture into a 2 L flask with a stir bar. Do not pour. Pipette up and down with the transfer pipettes to ensure homogenization before transfer.
3. When the entire deletion collection has been pooled, cover the 2 L flask with sterile foil and allow it to mix on a stir plate for 3 minutes to ensure homogenization of the pool.
4. The freshly harvested screening pool will not have a high enough cell density for the competition assay, and will have to be concentrated by centrifugation at $500 \times g$ before the frozen aliquots can be made. To concentrate the cell pool, use a centrifuge and large volume vessels. Concentrate the pool to an absorbance of 70 or greater at 600 nm by centrifuging at $500 \times g$ and decanting the excess media to adjust the OD. It is also advisable to confirm cells/mL using a haemocytometer. The recommended minimum density after concentration is 250 cells/strain/ μL . For a 5000 strain collection, that is 1.25×10^9 cells per mL. (*see Note 6*).
5. Once a high enough density is confirmed, add 1:1 v/v 30% sterile glycerol to the cell pool and gently homogenize via a stir plate. The glycerol dilutes the cell density, so it is important to have at least 250 cells/strain/ μL before adding the glycerol.
6. Aliquot the pooled deletion collection into freezer tubes. Aliquot volumes of 200-500 μL is best as only a small amount will be used in each experiment, and the pool cannot be used again once thawed. (*see Note 7*).
7. Store the aliquots at -80°C until use.

3.4 Pooled competition

The pooled competition is the point where the mutant pool is exposed to a compound, and this has 2 steps. First is determining the optimal dose of a drug for the assay. The chemical genomic assay is robust to dosing; however, signal can be maximized by dialing in the best inhibitory concentration. Using the pool as inoculum, find the final compound concentration that gives an inhibition level that gives a 20-50% reduction in growth compared to the solvent control after 24 h. Once the optimal dose is established, the next step is to perform the actual assay with replicates at the appropriate dose. It is important to include at least one control compound in the assay as well, a well-studied compound with known target. The readout of the control compound will let you know if the assay is working. Benomyl and MMS are good choices for control compounds. Benomyl targets tubulin and MMS damages DNA. A final concentration of 10-25 µg/mL of Benomyl or 0.01% MMS is an appropriate control dose. The chemical genomic profiles of these compounds can be used to assess success of the assay. In this step we describe how to determine the screening dose and then perform the pooled competition.

1. To determine the optimal dose, first thaw an aliquot of cells created in Section 3.3 and dilute the cells to the starting inoculum concentration (125-250 cells/strain/µL). The cells can be used directly once thawed and diluted.
2. Create cultures with 196 µL medium, 2 µL of a compound or solvent, and add 2 µL of the strain pool. Try to determine the compound and control compound dose that reduces growth of the pool by 20-50% relative to solvent control (although low/no inhibition can still be informative depending on performance of individual mutants). This is best accomplished using a dose curve.
3. Take a time zero (t0) measure of OD using a spectrophotometer, then incubate 24 h at 30 °C.
4. Measure the growth after 24 h and calculate the growth of the compound conditions relative to the solvent control. Plates can also be read continuously on an automated plate reader.
5. Find the dose of the compound that has growth 50-80% to that of the solvent, and this will be the screening dose to use in the next assay. For example, we have found a dose of 10-25 µg/mL of Benomyl or 0.01% MMS to be a good dose points for these control compounds.
6. Once dose is established, use this dose for all further experiments. Prepare wells for competition using 196 µL medium, 2 µL compound in solvent. It is best to have at least 4 replicates of each test compound or control compound. Run at least 4 solvent control conditions.

⁷The cell density of the screening pool stock, and thus the inocula used in experiments, is critical to obtaining informative chemical genomic signatures, as the ability to detect sensitive or resistant strains suffers when the number of cells per strain is either too high or too low. With too few strains there is not sufficient strain representation in the assay and functionally informative mutants may be missing; however, with too many cells, the signal of compound-mutant interactions is dampened as there is less capacity for growth. While the assay is somewhat robust to the starting cell density, we have found the optimal screening concentration to be 125-250 cells/strain. Higher densities are acceptable for the stock pool as it can be diluted prior to performing the growth assays, whereas it is more difficult to concentrate it if the density is too low.

7. Thaw another aliquot of cells created in Section 3.3 and dilute the cells to the starting inoculum concentration (125-250 cells/strain/ μ L). The cells can be used directly once thawed and diluted.
8. Add 2 μ L of the pooled cells to all wells containing the growth medium, mix and incubate at 30 °C for 48 h, recording the OD at 0 h, 24 h, 48 h. Plates can also be read continuously on an automated plate reader.
9. Harvest the cells after 48 h of growth by centrifugation and remove the supernatant, saving the pellet. Proceed to genomic DNA extraction (Section 3.5) or store a cell pellet at –80 °C until extraction. Make sure to remove all cells from the wells by pipetting up and down.

3.5 Genomic DNA extraction

After the pooled competition, the genomic DNA is extracted from the cells in preparation for amplification of the molecular barcodes. For less than 96 samples, it is preferred to perform individual genomic extractions. However, for larger scale projects, several 96-well genomic DNA kits and automated options are available, and we have found these to be comparable in quality (*see Note 4*).

Individual cultures: perform genomic DNA extractions on the 200 μ L cultures, scaling the kit specifications to the smaller volume. We have found that eluting the DNA with 35-50 μ L of elution buffer gives a good concentration for PCR.

96-well extractions: For 96-well extractions, both automated (*e.g.* Qiaextractor) and manual options exist. Many of these kits are not designed for yeast, and as such an extra cell wall digestion step is required. After pooled growth, harvest the cells by centrifugation and remove the supernatant. The cell pellets can be stored at –80 °C until needed. Before extraction, resuspend the cells in zymolyase solution and incubate for 1 h at 37 °C to digest the yeast cell wall. After this, proceed with the extraction according to kit specifications.

3.6 PCR and gel extraction

After the genomic DNA has been extracted, the next step is to PCR amplify the molecular barcodes that are used to identify members of the strain pool and assess their fitness in the presence of a compound. The yeast deletion collection has 2 molecular barcodes for each gene deletion, an “UPTAG” and a “DOWNTAG” (6). This method uses the “UPTAG” only.

It is at the PCR step that the index tags are added that allow multiplexed sequencing. Genomic DNA from each pooled competition will have an independent PCR reaction, each with a unique indexed primer plus the common primer. For 4 replicates, this means 4 separate PCR reactions each with a unique index primer and a common primer. These unique primers are designed with a 10 bp sequence that allows them to be pooled together (multiplexed) for sequencing and then de-multiplexed during analysis. A description of the index primer design and the resultant amplicon that will be sequenced can be found in **Fig 2a**. It is very important at this step to keep detailed notes on which indexed primer is matched with each experiment. For instance, the solvent control conditions may use primers 1-4, where the compound conditions use primers 5-8. The 10 bp index tag is what is used to tell

the analysis software how to demultiplex the data into the individual experiments. We have included an example of 12 indexed primers in **Table 1**, and a set of 96 unique indexed primers that we have assessed for performance within the software package and supporting material (available at www.github.com/csbio/barseq_counter). This step describes how to amplify the molecular barcodes with special indexed primers, pool, and then clean up the PCR product for barcode sequencing.

1. Purchased primers should first be diluted to working concentration using TE buffer. Prepare indexed PCR primers to a final concentration of 12.5 μ M for index primers and 100 μ M for the common U2 primer.
2. Reaction mixture (per reaction): 20.25 μ L of Taq mix, 0.25 μ L of U2 primer, 2 μ L of indexed primer, 2.5 μ L of genomic DNA. Always add genomic DNA last to avoid any chance of contaminating the primer stock.
3. PCR conditions: an initial denaturation of 5 min at 95 °C, followed by 30 cycles of 1 min at 95 °C, 30 s at 55 °C, 45 s at 68 °C, then a final extension time of 10 min at 68 °C.
4. Pool the PCR products from the individual PCR tubes for gel extraction by combining volumes of individual reaction mixtures into a single tube. For 8-24 samples, the entire volume of each reaction mixture can be pooled. For >24 samples, pool 10 μ L from each reaction.
5. Prepare a 2% agarose gel with Sybr Safe or Ethidium bromide for visualization. Using tape, make an extra large well that will accommodate the entire volume of the pooled PCR products. For >96 PCR reactions multiple gel extractions may be necessary.
6. Run the gel for 30-45 minutes under the following conditions: 120 V, 200 mA.
7. Identify the desired 267 bp PCR product band (*see Note 10*) by visualizing the gel under UV. There will be 3 bands: a lower primer band, a center band (267 bp, the desired product), and a higher band (>450 bp) (**Fig 2b**).
8. Carefully cut out the center band (267 bp) making sure not to cut any of the other bands. Try and excise a thin band, to minimize the amount of gel, and remove the slice to a tube for gel extraction.
9. Perform gel extraction on the excised gel. Elute the final product using a minimal buffer volume (25-50 μ L) to ensure a high template concentration.

3.7 Quantification and sequencing of the samples

Prior to sequencing, samples must be quantified before Illumina sequencing. Quantification is necessary to determine the correct amount of template DNA to load on the Illumina flow

¹⁰There will be three bands. A lower band ~100 bp that contains the unamplified primers, a middle band (267 bp) that contains the desired fragment, and a higher band (~500 bp) that results from non-specific amplification of the Illumina regions. Cut out only the middle band. As both the lower and upper band will contain Illumina regions, these will contaminate the flow cell if they make it to the sequencing reaction, and will result in fewer usable sequencing reads. The excised gel can be extracted following kit specifications; however, we recommend eluting the purified product with half the recommended volume to ensure a high concentration of PCR product.

cell, as with too much template the sequencing cluster density will be too high and will result in no data. If the cluster density is too low, the reads will not be sufficient enough to get an accurate estimate of strain fitness.

There are several approaches to quantifying the library for sequencing (*e.g.* Kapa qPCR kits, PhiX based qPCR, Bioanalyzer). We have found that the Kapa Illumina qPCR kits gives the most accurate estimation of the sample as it assesses only molecules amplifiable with the specific Illumina sequencing primers. A final template concentration of at least 10 nM is best to provide for sequencing. Even if a sequencing facility will quantify the samples prior to sequencing, provide quantification data when submitting the sample (*see Note 11*). In this step, we describe the method of sample quantification using the Kapa Illumina qPCR kit.

1. Prepare 3 replicates for each of the 6 standards.
2. Prepare 3 replicates of the following dilutions in 1X TE (elution buffer from the gel extraction kit) of the gel extracted sample: 1:2000, 1:5000, 1:10000.
3. Determine the concentration of each dilution via qPCR, using machine-specific software to calculate the concentrations.
4. Calculate the sample concentration using the Kapa Kit formula.
5. Run the samples on an Illumina platform (1 × 50 cycles) at a template concentration that will yield a cluster density of 700-900 k/mm². We have found a template concentration of 10-18 pM works best (*see Note 12*).
6. Obtain the fastq file after sequencing

3.8 Data analysis

Following sequencing, it is finally time to get a glimpse of the data. To generate a chemical genomic profile, the raw sequence data must be processed, de-multiplexed based on the index tags, and finally a chemical genomic score assigned to each mutant in the different compound conditions. We have provided a set of open source Python and R-package scripts that can be used to perform these analyses. The script package entitled “barseq counter” is available from www.github.com/csbio/barseq_counter. Within this package are all the required instructions, scripts, and tools necessary to process, demultiplex, count, and detect compound-mutant interactions. We have also provided a sample dataset (fastq files plus

¹¹There are several methods to quantify samples for Illumina sequencing. Sequencing facilities often have their preferred methods, and it may be best to follow their recommendations. We have found that the Kapa Illumina qPCR kits to be very reliable in estimating product, as these rely specifically on the Illumina regions to amplify, and detect the quantity of only amplifiable DNA. qPCR data can be paired with Qbit and Bioanalyzer data to get high confidence quantification. We recommend for qPCR, that all samples be run in triplicate, control curves and dilutions of the samples. We have found using dilutions of 1:2000, 1:5000, and 1:10000 are best for getting accurate quantification. 1:1000 dilutions are often still too concentrated.

¹²Our method has been optimized for Illumina sequencing; however, it could be adapted to any next-generation sequencing platform. Smith et. al. (2010) describes both Illumina and Solexa barcode sequencing (15). As these are single-end reads, we can use a higher-cluster density on the Illumina flow cell, and a higher template concentration. We have found that, for experiments using the 5000 strain pool, a running concentration of 10-18 pM as determined by the Kapa qPCR kit results in an optimal quality and quantity of sequencing reads. However, it is best to work with the sequencing facility operators when dialing in the running concentration. For the 5000 strain collection and the MiSeq platform, up to 25 samples can be pooled and analyzed in one MiSeq run. For HiSeq 2500 Rapid Run, up to 192 samples can be pooled and analyzed in a HiSeq 2500 Rapid Run (2 flowcells, 1 sample). For single lane HiSeq 2000/2500, up to 96 samples can be pooled and analyzed in each lane of a HiSeq 2000/2500 run (768 samples per full 8-lane sequencing run). We do not recommend using PhiX in the lane with the samples, as it reduced read counts and does not help read quality.

decode and control input files) with detailed instructions on using the scripts so that a first-time user can learn how to operate the script package and get a feel for what the output data will look like. The sample dataset is hosted at http://lovelace-umh.cs.umn.edu/chemical_genomics_tools/barseq_counter. “Barseq counter” requires substantial computational power, and is best run from a server with the latest versions of Python and the R-package installed. This software will pre-process the raw fastq sequence file, then demultiplex based on the known indexed primers list in the “decode” file that the user provides. Then, it will generate a count matrix of the data with the specific sequence counts for each mutant under each experimental condition.

The software will also provide basic quality control (QC) steps to help assess the overall success of the sequencing run. It will tell the percent of sequence reads passing the filter, the number of reads for each index tag and the number of reads for each mutant. This QC output is important; if a particular index tag has abnormally low counts then the data from that tag may not be useful. In general, if the sequencing read has an average of 100 counts per strain, then it can be considered successful. Of course some strains will have either low counts or high counts based on their response to the compounds, but the average count across all mutants is what to look for to determine how well the sequencing run performed.

The final step uses runEdgeR.R, an R script wrapper around the EdgeR package, for determining the differential growth and statistical significance for each mutant relative to the control conditions. The runEdgeR.R script normalizes the data against the solvent control conditions, and allows the detection of drug-gene interactions. The use of EdgeR in barcode sequence experiments is described well in Robinson et al 2014 (11, 12), where the list of fold changes is the chemical genomic profile, and the genes in this list give functional insight into the compound's mode of action and potential cellular target. The R script also generates box and correlation plots of the data to further assess data quality and determine the agreement between replicates. This section describes how to install, use, and interpret chemical genomic data.

1. Download and copy the folder ‘barseq_counter’ to a computer that will run Python and the R-package for processing the data. This is a resource intensive process, and is best run from a server rather than a PC. Within the “Barseq_counter” package is a program called Agrep 3.14, which is used to count the barcodes and must be installed if not already present. The file folder “barcodes” contains the mapping of the strain specific barcodes to their ORF name, and the file “allupbarcodes.txt” contains the most up to date list.
2. Obtain a fastq file after sequencing and create a main folder for the experiment (e.g. Chemgen1). We have provided a test dataset here http://lovelace-umh.cs.umn.edu/chemical_genomics_tools/barseq_counter. Within this folder create a sub-folder entitled ‘data’. Place the fastq file in the data folder. If there are multiple fastq files, these all can be placed in the “data” folder and the scripts will concatenate them during processing.
3. The first processing step is to remove the excess sequence data from the reads (e.g. the common priming region). This is accomplished with the script

“preprocess_MIseq_10bp.py.” Within Python, navigate to the experiment folder (not the “data” folder) and enter:

```
python barseq_counter/scripts/preprocess_MIseq_10bp.py data barseq.txt
```

This command will unite them into a single processed output file “barseq.txt”. The script removes the common priming region of the sequencing read, which is not used in analysis. Further, the script separates the 10 bp index tag from the gene barcode.

4. The next step is to demultiplex the data based on the unique index tags and count the strain specific barcodes. This step requires a user input file that tells the software which index tag corresponds to each experiment. This file is the index tag decode and is a text file with the index tag followed by the experimental condition and has the following format:

GATTAGCCTC	DMSO
AATGAGCCGT	DMSO
ACGCGGATTA	DMSO
GCTTACGGAA	MMS
CGGTAGACTA	MMS
ATTGCCGGAA	MMS
GACATGCTAG	Benomyl
TACGCTGCAT	Benomyl
GTCAAGCACT	Benomyl

See the example file entitled “decode.txt” as an example. Do not use special characters in the compound names (*e.g.* %, *), and keep a detailed file with all dose points for cross reference. Make sure all replicates’ names are the same and not indicated by a replicate number or letter in the decode file (*e.g.* do not use BenomylA, BenomylB, etc), as it is necessary that they are distinguished only by the index tag and maintain the same naming scheme for downstream processing. To demultiplex the data, have the decode file in the experiment folder and use the command:

```
python Barseq_counter/scripts/processBarSeq_rd.py barseq.txt decode.txt
Barseq_counter/barcodes/allupbarcodes.txt barseq.processed.cerevisiae.txt
```

The output of this step is “barseq.processed.cerevisiae.txt,” which contains a matrix of individual strain counts across all experiments. In this matrix, the gene names of the mutants are listed as the systematic ORF. We have provided a script that will convert the ORF name to the common yeast name. This script is called “convertORF2common.py”. To run this use the follow command from within the experiment folder:

```
python barseq_counter/scripts/convertORF2common.py
barseq.processed.cerevisiae.txt barseq.processed.common.txt
```

The file `barseq.processed.cerevisiae.txt` is the count matrix for the experiment and will be used to determine chemical genetic interactions.

5. Before moving on to determine chemical genetic interactions, it is important to assess sequence quality, to make sure there are enough sequence reads and spot any potential problems with the data. First make a new folder for the quality reports. While still in Python, use the command:

```
"mkdir -p <folder name>_reports"
```

to create a folder for the read quality reports. Next use the command:

```
python Barseq_counter/scripts/generateReport_MIseq.py data barseq.txt
barseq.processed.cerevisiae.txt <folder name>_reports/cerevisiae
```

This command will generate distribution plots of the index tag and barcode counts, in addition to text file summaries. Here the average sequencing counts for each index tag and barcode can be assessed. These data can be used to estimate the counts per strain for each indexed conditions by dividing the total counts for an index tag by the number of strains, and ideally there will be >100 counts per strain.

6. The raw count matrix from Section 3.7. step 4 is used to determine chemical genetic interactions. This is done using the R-package "EdgeR" and a user input file that identifies the control (solvent conditions). EdgeR normalizes the count data for each experimental condition and estimates the differential growth against the control conditions to generate a fold change for each mutant in the presence of a compound. A fold change of >1 indicates increased growth, whereas <1 indicates reduced growth or sensitivity of the mutant to a compound compared to the solvent control. EdgeR also generates an adjusted P-value for responsive strains, which is a measure of statistical significance of the fold-change, which has been corrected for multiple comparisons. The control conditions file is a text file list of index tags associated with the solvent conditions using the following format:

```
DMSO_GATTAGCCTC
```

```
DMSO_AATGAGCCGT
```

```
DMSO_ACGCGGATTA
```

To run EdgeR, invoke R while in the experiment folder and use the command:

```
runEdgeR.R --threshold 10 barseq.processed.cerevisiae.txt controls.txt
```

The threshold is set to make sure EdgeR ignores conditions with very low read counts; we generally set the threshold at 10 counts. In the 'csv' output folder, EdgeR will provide a list of compound responsive strains for each compound sorted based on the significance of their fitness change relative to the solvent control. The output "pdf" and "png" folders from EdgeR provide heatmap plots of the data to assess replicate agreement and global distribution of sensitive and resistant mutants as volcano plots. This is the starting point to interpret the chemical genomic data and make predictions of compound mode of action. Sort the list by fold change, and start by searching for functional enrichment among the top

10-20 sensitive or resistant strains (<http://go.princeton.edu/cgi-bin/GOTermFinder>). Use the control compound conditions to assess assay success. For example for Benomyl, within the top sensitive strains are gene mutants in tubulin related processes (e.g. CIN1, GIM4, PAC2). For MMS, there is significant enrichment for mutants involved in DNA repair (e.g. PSY3, SAE2, RAD4). Look for functional enrichment in the sensitive and resistant mutants of the unknown compounds to gain functional insight into what they may be targeting.

This chapter describes the basic steps for generating chemical genomic profiles using the non-essential yeast deletion collection and barcode sequencing. The profile can be used to unveil the cellular target or mode of action of novel compounds. This approach can be repeated using the heterozygous yeast deletion collection of essential genes to cover greater target space. Further, the chemical genomic profile can be correlated with the genetic interaction profiles of the yeast genetic interaction network and give further insights into mechanism (4, 10). Chemical genomics paired with barcode sequencing can provide an unbiased, high-throughput screening method of rapidly linking compounds to their cellular targets.

Acknowledgements

JP, SM, IO are funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494. SC is supported by a RIKEN Foreign Postdoctoral Researcher Award. CM, SWS, and RD are supported by grants from the National Institutes of Health (1R01HG005084-01A1, 1R01GM104975-01, R01HG005853), a grant from the National Science Foundation (DBI 0953881), and by the CIFAR Genetic Networks Program. SWS is supported by an NIH Biotechnology training grant (5T32GM008347-22). CB is supported by the Canadian Institutes of Health Research (CIHR MOP-57830). RA is supported by the Natural Science and Engineering Research Council (NSERC) of Canada, the Canada Foundation for Innovation (CFI), and the Canadian Cancer Society Research Institute.

References

1. Parsons A, et al. Exploring the Mode-of-Action of Bioactive Compounds by Chemical-Genetic Profiling in Yeast. *Cell*. 2006; 126:611–625. [PubMed: 16901791]
2. Ho CH, et al. Combining functional genomics and chemical biology to identify targets of bioactive compounds. *Curr Opin Chem Biol*. 2011; 15:66–78. [PubMed: 21093351]
3. Fung S-Y, et al. Unbiased Screening of Marine Sponge Extracts for Anti-Inflammatory Agents Combined with Chemical Genomics Identifies Girolline as an Inhibitor of Protein Synthesis. *ACS Chem Biol*. 2013; 9(1):247–57. [PubMed: 24117378]
4. Williams DE, et al. Padanamides A and B, Highly Modified Linear Tetrapeptides Produced in Culture by a *Streptomyces* sp. Isolated from a Marine Sediment. *Org Lett*. 2011; 13:3936–3939. [PubMed: 21749075]
5. Giaever G, et al. Chemogenomic profiling: Identifying the functional interactions of small molecules in yeast. *Proc Natl Acad Sci U S A*. 2004; 101:793–798. [PubMed: 14718668]
6. Giaever G, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*. 2002; 418:387–391. [PubMed: 12140549]
7. Kim D-U, et al. Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nat Biotech*. 2010; 28:617–623.
8. Kitagawa M, et al. Complete set of ORF clones of *Escherichia coli* ASKA library (A Complete Set of *E. coli* K-12 ORF Archive): Unique Resources for Biological Research. *DNA Res*. 2006; 12:291–299. [PubMed: 16769691]
9. Smith AM, et al. Quantitative phenotyping via deep barcode sequencing. *Genome Res*. 2009; 19:1836–1842. [PubMed: 19622793]

10. Costanzo M, et al. The Genetic Landscape of a Cell. *Science*. 2010; 327:425–431. [PubMed: 20093466]
11. Robinson DG, Chen W, Storey JD, Gresham D. Design and Analysis of Bar-seq Experiments. *G3 GenesGenomesGenetics*. 2014; 4:11–18.
12. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma Oxf Engl*. 2010; 26:139–140.
13. Pierce SE, Davis RW, Nislow C, Giaever G. Genome-wide analysis of barcoded *Saccharomyces cerevisiae* gene-deletion mutants in pooled cultures. *Nat Protoc*. 2007; 2:2958–2974. [PubMed: 18007632]
14. Andrusiak, K. [April 24, 2014] Adapting *S. cerevisiae* Chemical Genomics for Identifying the Modes of Action of Natural Compounds.. Thesis. 2012. Available at: <https://tspace.library.utoronto.ca/handle/1807/32456>
15. Smith AM, et al. Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucl Acids Res*. 2010:gkq368.

Figure 1.
Overview of the steps of chemical genomic profiling by barcode sequencing

Figure 2.

PCR amplicon design and gel extraction of the amplified barcodes. The indexed primer design and components of the PCR amplicon used in sequencing (**A**). Using a 2 % agarose gel with Syber Safe of Ethidium Bromide, run the pooled PCR product for 45 min using a 1 kb ladder for reference and visualize under UV. Three bands will be apparent, a lower band of unused primers, the middle band at 267 bp, and an upper band of amplification artifacts. Excise the center band for gel extraction (**B**). The other bands contain the Illumina regions of the primers and if run on the sequencing flowcell, will form clusters but will not provide usable reads (loss of read depth).

Figure 3.

Distribution of Barcodes based on pooling method. We constructed 2 pools of 300 strains using 2 different pooling methods: agar scraping and mixing of liquid cultures. We then grew the pools in our chemical genomics assay. While the pools performed similarly, we found that the liquid pool had better distribution of strains, as determined by fraction of individual strains in the pool. In this figure, a straighter line indicates a more even strain distribution.

Table 1

Example set of 12 indexed primers and their 10 bp index tags. The entire primer contains the Illumina specific region, the index tag, the common priming site for barcode amplification. The common reverse primer contains an Illumina specific region and a common priming site form the KanMX gene region of the deletion insert.

Index Tag	Entire indexed primer (5' - 3')
AATAGGCGCT	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTAATAGGCGCTGATGTCCACGAGGTCCTCT
TACAGTTGCG	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTTACAGTTGCGGATGTCCACGAGGTCCTCT
ATCCTAGCAG	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTATCCTAGCAGGATGTCCACGAGGTCCTCT
GATTAGCCTC	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTGATTAGCCTCGATGTCCACGAGGTCCTCT
AATGAGCCGT	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTAATGAGCCGTGATGTCCACGAGGTCCTCT
ACGCGGATTA	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTACGCGGATTAGATGTCCACGAGGTCCTCT
GCTTACGGAA	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTTACGGAAAGATGTCCACGAGGTCCTCT
CGGTAGACTA	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTCGGTAGACTAGATGTCCACGAGGTCCTCT
ATTGCCGGAA	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTATTGCCGGAAAGATGTCCACGAGGTCCTCT
GACATGCTAG	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTGACATGCTAGGATGTCCACGAGGTCCTCT
TACGCTGCAT	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTTACGCTGCATGATGTCCACGAGGTCCTCT
GTCAAGCACT	AATGATACGGCGACCCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTGTCAAGCACTGATGTCCACGAGGTCCTCT
Common reverse primer	CAAGCAGAAAGACGGCATACGAGCTCTTCCGATCTGCACGCTCAAGACTGTCAAGG