HAL
open science

Identification of Links Between Cellular Pathways by Genetic Interaction Mapping (GIM).

Christophe Malabat, Cosmin Saveanu

To cite this version:

Christophe Malabat, Cosmin Saveanu. Identification of Links Between Cellular Pathways by Genetic Interaction Mapping (GIM).. Frederic Devaux. Methods in Molecular Biology (1064-3745), 1361, Humana Press, pp.325-43, 2016, Yeast Functional Genomics: Methods and Protocols, 978-1-4939-3078-4. 10.1007/978-1-4939-3079-1_18 . pasteur-01427613

HAL Id: pasteur-01427613

https://pasteur.hal.science/pasteur-01427613

Submitted on 5 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Identification of links between cellular pathways by Genetic Interaction Mapping (GIM)

Christophe Malabat and Cosmin Saveanu

Génétique des Interactions Macromoléculaires (UMR3525-CNRS)

Institut Pasteur, 25 rue du Docteur ROUX, 75015 PARIS

Corresponding author: cosmin.saveanu@pasteur.fr

Summary

The yeast systematic deletion collection offered the basis for a number of different strategies that establish functional links between genes by analyzing the phenotype of cells that combine two different deletions or mutations. A distinguishing feature of the collection is the presence of molecular barcodes at each deleted locus, which can be used to quantify the presence and abundance of cells bearing a given allele in a complex mix. As a result, a large number of mutants can be tested in batch cultures, replacing tedious manipulation of thousands of individual strains with a barcode microarray readout. Barcode-based genetic screens like Genetic Interaction Mapping (GIM) thus require little investment in terms of specific equipment, are fast to perform, and allow precise measurements of double mutant growth rates for both aggravating (synthetic sick) and alleviating (epistatic) effects. We describe here protocols for preparing the pools of haploid double mutant *S. cerevisiae* cells, testing their composition with barcode microarrays, and analyzing the results to extract useful functional information.

Key words: Genetic screen, Gene deletion, Growth rate, Barcode microarray, Data analysis.

1 Introduction

Tests of the effect on growth rate for a large number of combinations of gene deletions became possible in the last decade through technical advances that arose after the completion of *S. cerevisiae* genome sequencing. An essential development in this respect has been the creation of a collection of systematic deletion for all the predicted yeast genes ***(1, 2)***. Each cell of this collection bears specific short 20 nucleotide long sequences that uniquely identify the affected locus. These "molecular barcodes" allow the analysis of thousands of strains in parallel, to estimate growth of any single strain in a complex mixed population. The second most important advance in the field has been the invention of methods that would allow efficient production of thousands of double mutant strains. Two strategies were employed to this end. One, called SLAM (Synthetic Lethality Analysis with Microarrays), was based on massive transformation of cells from the collection with DNA fragments that ensure the replacement of a gene of interest with a selection cassette ***(3, 4)***. The other, which was employed in several laboratories under various names, is based on efficient mass mating ***(5)*** of haploid yeast cells, bringing together two mutations of interest. The development of a method for efficient selection of double mutant haploid strains issued through meiosis from these diploids was crucial to the ability to obtain and test a large number of mutation combinations in parallel. Both Synthetic Genetic Array analysis ***(SGA, 6, 7)*** and Genetic Interaction Mapping ***(GIM, 8)*** use this approach. Quantitative tests of the extent by which two concomitant mutations affect growth rate in comparison with each single mutation present in a yeast strain, were pioneered in Epistatic Miniarray (E-MAP) studies ***(9)***, and used in SGA ***(10)*** and GIM ***(8)*** screens. The GIM method,

presented here, combines the efficient generation of haploid double mutant strains through mating and sporulation with a direct quantitative test of strains growth based on DNA microarrays.

Genetic screens of the type described above can be extended to other collections of mutant strains. For example, substituting the normal 3' UTR region of a gene with a long 3' UTR destabilize mRNA through nonsense-mediated mRNA decay ***(NMD, 11)***. The reduction of the amounts of proteins allowed testing of phenotypes for some of the yeast essential genes ***(9)***. Our recent analysis of a collection of such strains indicate, however, that this strategy is biased and cannot guarantee that the obtained strains show a phenotype, since NMD shows preference for mRNAs with short coding sequences ***(12)***. The study of phenotypes for essential gene mutants can also be done with collections of thermosensitive yeast strains ***(10)***.

The details provided in the present chapter should be helpful to set up and perform reproducible GIM screens and identify either direct genetic interactions or specific interaction profiles that indicate which cellular pathways are required for the cells to adapt to the absence of a given gene. Our experience with large scale quantitative data indicate that crucial to the identification of patterns or interesting correlation is easy visual data representation. Several software solutions answer this question including Mayday ***(13)***, geWorkBench ***(14)***, or Perseus ***(15)***.

In the end, GIM screen results are only as good as the capacity of the experimenter to deduce useful information from them. In most of the cases, the screens are only the starting point for projects that need years of focused work until completion. Thus, such screens are most of the time a discovery tool that

provides first hints for focused analyses in functional genomics.

The preparation of the query strain as well as the relative timing of a GIM screen is indicated in **Fig. 1**. We perform between 4 and 16 parallel screens in parallel and the limiting step is the selection and culture of the pool of haploid double mutants.

2 Materials

2.1 Growth media, antibiotics and molecular biology reagents

1. YPD medium: 2% D-glucose, 1% Difco yeast extract, 1% Difco Bactopeptone. Autoclave at 110°C for 20 minutes.
2. GNA medium: 5% D-glucose, 3% Difco nutrient broth, 1% Difco yeast extract. Autoclave at 110°C for 20 minutes. Prepare fresh before use.
3. Sporulation medium: 1% potassium acetate (10% stock), 0.005% zinc acetate (0.5% stock), uracile (2 mg/100ml, 0.2% stock), histidine (2 mg/100ml, 0.4% stock) and leucine (6 mg/100 ml, 1.2% stock in dH_2O). Resuspend the stocks in dH_2O and filter sterilize.
4. Antibiotics: 0.2 mg/ml Hygromycin B (stock 50 mg/ml); 20 µg/ml Nourseothricin, clonNat (Werner Bioagents, ref 5.1000, 200 mg/ml stock); 0.2 mg/ml G418 sulphate (100 mg/ml stock).
5. Synthetic complete medium without uracile: 6.7 g/L Yeast Nitrogen base without aminoacids, with ammonium sulfate, 2% glucose and 0.2% amino acids mix. The mix of amino acids is composed of alanine, arginine, aspartic acid, asparagine, cysteine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine,

tryptophane, tyrosine and valine, in equal weight proportion, with the exception of leucine, for which twice as much is added. In the same mix, add for 1 g of amino acid 0.5 g of adenine. Autoclave or filter sterilize. For plates 2% agar is used in the final medium composition.

6. Culture plates (96 deep wells, 2 ml), 96 long pins replicator.

7. Gaz permeable and aluminium seal for 96 well plates.

8. NEB4 buffer: 50 mM potassium acetate, 20 mM Tris-acetate, pH 7.9 at 25°C, 10 mM magnesium acetate, 1 mM dithiothreitol.

9. *Asc*I and *EcoR*I restriction enzymes. pGID3 plasmid.

2.2 Microarray reagents

1. Oligonucleotides: **U1** 5'-GAT GTC CAC GAG GTC TCT; **KU** 5'-AAG AAG AAC CTC AGT GGC; **D1** 5'-CGG TGT CGG TCT CGT AG; **KD** 5'-GGA TCT TGC CAT CCT ATG; **U2block** 5'-CGT ACG CTG CAG GTC GAC; **D2block** 5'-ATC GAT GAA TTC GAG CTC; **U2-Cy3/5** 5'-Cy3/5-GTC GAC CTG CAG CGT ACG; **D2-Cy3/5** 5'-Cy3/5-CGA GCT CGA ATT CAT CGA T;

2. Custom Agilent microarray (8x15k version for barcodes, Agilent-026035 Scer_barcode_v2_200911, platform GPL18088 at the GEO database, www.ncbi.nlm.nih.gov/geo/) (*see* **Note 1**).

3. Extraction buffer: 2% Triton X100, 1% SDS, 0.1M NaCl, 10 mM Tris-Hcl pH 8, 1 mM EDTA.

4. TE buffer: 10 mM TrisHCl pH8, 1 mM EDTA.

5. Acid washed glass beads of 0.4-0.6 mm diameter.

6. Phenol-chloroform-isoamyl alcohol mix (25:24:1, pH 8).

7. Linear acrylamide, 5 mg/ml.

8. DIG Easy Hyb solution (Roche Applied Science 1.603.558).

9. 20X SSPE buffer: 3 M NaCl, 20 mM EDTA, 0.2M NaH_2PO_4, pH 7.4.

10. Agilent hybridization chamber and gasket slides for 8x15k microarrays.

11. Agilent stabilization and drying solution, required to protect Cy5 fluorescence from fading under the action of ambient O_3.

12. Scanner (GenePix 4000B) and scanning software (GenePix Pro 6, Molecular Devices, LLC).

13. System for normalization and data analysis (R from cran.r-project.org/ with the ggplot2 package installed).

3 Methods

3.1 Obtention of a pool of *MAT***a** mutant cells from the collection

This is probably one of the most time consuming steps when initiating the first series of GIM screens in a laboratory. It requires manipulation of 74 plates with 96 wells under sterile conditions (*see* **Note 2**). The pools of mutants can be obtained in large amounts and aliquoted, to be used by several laboratories, as a common resource.

1. Thaw several plates (usually no more than 12, for ease of manipulation) from the collection on the bench until the medium is half-molten. Carefully pick cells from the plate by using a 96 pin replicator. Pay attention that no droplets of medium are transported from one well to another. Transfer the cells to two different 96 well plates that are already filled with YPD medium. A standard 96 well plate will serve as a duplicate of the original collection, while the

second one, with deep wells that contain up to 2 ml of YPD, will constitute the culture of individual strains to constitute the pool.

2. Add a fresh aluminium sticking cover on the original plate and return it to -80°C (*see* **Note 3**).

3. Leave the cells in the new plates to grow for 48-72 hours at 30°C, under a gaz permeable seal. No continuous agitation is required.

4. Verify that the cells have grown in all the wells that contained cells in the original plate and are recorded as such in the spreadsheet that accompanies the collection. Transfer the cells to a single flask by using a multichannel pipetting device. Vortex briefly, centrifuge at 3200 g for 10 minutes at room temperature. Resuspend the cells in YPD containing 25% (w/v) glycerol and take an OD_{600nm} reading. Store the pool of cells in aliquots of different sizes that will be directly used in the GIM screens.

3.2 Generation of a “query” mutant strain for the screen

1. Thaw one or several haploid BY4742 strains from the collection that carry the G418 resistance cassette as a replacement of the coding sequence for the genes of interest. Carry out the maximum number of validations required to be sure the strain lacks the gene of interest and shows no additional mutations (*see* **Note 4**).

2. Prepare enough pGID3 plasmid to be able to transform as many strains as you wish (1 µg digested plasmid/strain). For a single strain, place 1 µg pGID3 plasmid in 10 µl of NEB4 buffer containing 5 U of *Asc*I and 5 U of *EcoR*I. Digest for 1 hour at 37°C. Test digestion efficiency by electrophoresis on 1 % agarose gel. Two bands corresponding to the digestion fragments (2066 nt

and 2484 nt) should be obvious.

3. Prepare for each strain 10 ml of culture in YPD medium, leave to grow overnight. In the next morning, the culture should be saturated (OD_{600nm} between 10 and 15). Dilute culture in 50 ml fresh YPD to an OD_{600nm} of 0.15. Proceed with a DMSO enhanced yeast transformation protocol ***(16)*** with 1 µg of digested pGID3. Spread transformed cells on YPD plates containing nurseothricin to select for the recombination event. Typical yields are 50 to 500 colonies, visible after 2 days of plate incubation at 30°C. Streak single colony cells on YPD plates containing G418 and on YPD plates containing nurseothricin. Store cells that have lost G418 resistance and became resistant to nurseothricin at -80°C.

3.3 Obtention of a pool of double mutant strains

1. Transform previously obtained strains, that bear the pMFα2-Nat^R cassette as a replacement for the coding sequence of the gene of interest, with 0.3 µg pGID1 plasmid, containing two different markers: URA3 and Hygromycin resistance (*see* **NOTE 5**). Select transformants on SC-URA plates (optionally, these strains can be also stored at -80°C, to speed up the time until a screen starts).

2. The evening before starting the screen, prepare a pre-culture of the query strain in 10 ml YPD medium containing hygromycin. The next morning, dilute the cells to an OD_{600nm} of 0.2 in 25 ml GNA + hygromycin medium and leave to grow until the OD_{600nm} reaches 0.8. Thaw an aliquot of the pool of mutants in 25 ml GNA medium and incubate for 30 minutes at 30°C with agitation; the expected OD_{600nm} is 0.4 (*see* **Note 6**).

3. Combine the two cultures in a 50 ml tube, centrifuge 10 min. at 3200 g at

room temperature, resuspend the cells in 0.5 ml GNA medium and spread the suspension on a 90 mm diameter GNA plate. Leave at 30°C for 5 hours to allow mating.

4. Recover (scrape) the cells from the plate in 1 ml GNA medium containing hygromycin and G418 and resuspend them in 100 ml of the same medium. Incubate for 18 hours, usually overnight, at 30°C with agitation.

5. Take an OD_{600nm} reading for the overnight culture (dilute 1/50 in GNA medium) and recover an amount of culture that, when diluted in 150 ml would yield an OD_{600nm} of 0.8 (typically 6-10 ml). Centrifuge at room temperature in a 50 ml Falcon tube for 10 minutes at 3 200 g.

6. Thoroughly resuspend the pellet in 25 ml sporulation medium. Centrifuge a second time and resuspend the cells in another 25 ml sporulation medium (*see* **Note 7**). Resuspend the cells in 150 ml sporulation medium in a 1 liter culture flask.

7. Leave the cells at 25°C for 5 days with continuous agitation and shift the cultures to 30°C for an additional 3 days. At the end of this sporulation period check by light microscopy that specific tetrads are a sizable proportion of the cells.

8. Recover the cells from 100 ml of sporulation flask by centrifugation at 3200 g, 10 minutes, RT. Resuspend the pellet in 1 l YPD in a large 2 liter culture flask and incubate with agitation at 30°C for 5.5 hours.

9. Add antibiotics (G418 and nourseothricin) to the YPD medium and the cells and leave with agitation at 30°C.

10. Take repeated readings of OD_{600nm} until the values reach a value of 2. Dilute

the cells in 1 l of fresh YPD medium with antibiotics (G418 and nurseothricin) to a DO_{600nm} of 0.1. Estimate the number of generations from the slope of the growth curve of the newly diluted culture to know at which moment the average number of generations is 18 (in general 45-60 hours, *see* **Note 8**).

11. Recover the equivalent of 10 OD_{600nm} culture for each screen and the corresponding reference pool of cells. Pellet the cells by centrifugation at 3200 g, 10 min, RT, and resuspend in 1 ml ice cold H_20, then split in two aliquots in 1.5 ml Eppendorf tubes. Pellet the cells by 20 seconds of 13 000 g centrifugation at 4°C, remove supernatant and freeze the cells at -80°C until further processing.

3.4 Growth speed estimation from barcode microarray data

1. Resuspend the cell pellet recovered in the previous step in 0.2 ml extraction buffer and add approximately 0.2 ml acid-washed glass beads (0.4-0.6 mm diameter, Sigma) and 0.2 ml phenol-chloroform-isoamyl alcohol mix (25:24:1, pH 8). Vortex vigorously for 7 minutes under a chemical hood.

2. Add 0.2 ml TE buffer and vortex for 5 seconds. Centrifuge at 13 000 g for 10 minutes, at RT. Transfer the supernatant in a 1.5 ml Eppendorf tube that already contains 1 ml pure ethanol. Mix by turning the tube upside down several times.

3. Centrifuge at 13 000 g, at RT, for 5 minutes. Wash the white small pellet that sticks to the bottom of the tube with 0.5 ml of 70% ethanol. Centrifuge at 13 000 g for 5 minutes at RT. Throw the supernatant, leave the pellet at room temperature to dry out, between 5 and 10 minutes and resuspend the dry pellet in 30 µl of TE (optionally add 10 ng/ml RNase A).

4. Use 1 µl of the previously prepared DNA extract as template in each of two PCR reactions that are set up using oligonucleotides U1-KU (uptag PCR 1, UPCR1) and D1-KD (downtag PCR 1, DPCR1) (*see* **Fig. 1a** and **Note 9**). Set up standard 50 µl reactions for each PCR reaction (25 cycles, 94°C denaturation 30 sec., 50°C annealing 30 sec., 72°C elongation 30 sec). Verify the size and abundance of PCR products by electrophoresis on a 1% agarose gel.

5. Prepare four different PCR reactions for each comparison between a screen using a query mutation and a screen performed under identical conditions with a reference mutation (*see* **Note 10**). Set up standard 50 µl PCR reaction with 0.5 µl UPCR1 or DPCR1 in the following combinations with oligonucleotides: UPCR1 query with oligonucleotides U1 and U2-Cy3, UPCR1 reference with oligonucleotides U1 and U2-Cy5, DPCR1 query with oligonucleotides D1 and D2-Cy3, DPCR1 reference with oligonucleotides D1 and D2-Cy5. Use 15 cycles of amplification (94°C 15 sec, 55°C 15 sec, 72°C 15 sec). Test the reactions by electrophoresis using a 3% agarose gel and a loading buffer that only contains xylene cyanol. The expected size of PCR products is 60 nt.

6. Mix the four PCR products obtained in the previous step. Add a pre-mix containing 5µl each of oligonucleotides U1, D1, U2block and D2block (100 µM each, *see* **Note 11**) and 2 µl linear poly-acrylamide (5 mg/ml), as carrier. Add 22.5 µl sodium acetate (3M, pH=5.2), mix well and precipitate with 550 µl pure ethanol. Leave at -20°C for at least one hour.

7. Centrifuge 30 minutes at 15 000 g, 4°C. Pay attention to the purple pellet when removing the supernatant. From now on, avoid direct exposure of the

pellet or solution to light. Wash the pellet once with 70% ethanol. Recentrifuge 5 minutes at 15 000 g, 4°C. Remove the ethanol and leave the tube on the bench for 5 minutes to dry.

8. Add 50 µl DIG Easy Hyb buffer to the pellet. Mix by repeated pipetting. Dilute 20 µl of the sample with 100 µl DIG Easy Hyb and keep the rest at -20°C, for the case when the hybridization needs to be repeated. Heat for 2 minutes at 95°C and then switch to ice for 5 minutes. Keep the sample at room temperature until hybridization.

9. Add the cover slide, with the gaskets, to the lower part of the hybridization chamber. Pipet carefully 47 µl of sample in the middle of a gasket square. Carefully layer the microarray slide over the sample, oligonucleotide side down (can be detected by breathing on the sides of the slide) and close the hybridizaton chamber. Pay attention that a single air bubble is formed in the chamber. Place the chamber on a rotating wheel (*see* **Note 12**) in a 24 °C incubator for at least 12 hours.

10. Prepare several dilutions of SSPE buffer (6x, 2x, 0.2x) containing 0.05% Triton X100. Pre-heat 50 ml SSPE 6X, 0.05% Triton X100 at 30°C. Open the chamber and transfer the two slides together in a Falcon tube with the wash 6X SSPE buffer. Remove the slide with the gaskets and transfer the microarray slide to a second bath of 6x SSPE, 0.05% Triton X100. Leave for 5 minutes with occasional agitation, then move the slide to the 2X SSPE solution, 0.2X SSPE solution and finally to a 0.2X SSPE solution containing no Triton X100. Incubate at each step for 5 minutes (RT).

11. After the last wash, slowly remove the slide from the liquid, dry its lower edge on a piece of Whatman paper, and plunge in stabilization solution for 10

seconds (under a chemical hood). Slowly remove the slide from the liquid so that no droplets remain; droplets can also be removed by a flux of dry air.

12. Scan the slide on a GenePix 4000B or equivalent scanner that is capable of 5 µm resolution. Analyze the scan images using an associated file that describes the position and annotation of each spot (**gal** file for GenePix Pro 6 or 7). The results are exported in the **gpr** format, a text file that contains columns separated by the **tab** character.

3.5 Microarray data analysis

1. We provide here an example of R session commands used to normalize data post-acquisition and to be able to do diagnostic plots. The input text file is the **gpr** result of image analysis with GenePix Pro.

2. Define a function allowing to extract the ORF names from the name of the probes (to be changed depending on how probes are identified). An example of identifier is YDR439W-U, where U shows that the barcode is the one located in the upstream position of the KanMX4 cassette. The "**namesextract**" function uses regular expressions to identify and grab ORF information from the identifier.

```
namesextract <- function(IDs) {
    orfnms1 <- gsub("(Y......-[ABC]?).*", "\\1", IDs)
    orfnms2 <- gsub("-$", "", orfnms1)
    return(orfnms2)
}
```

3. Read the tab delimited file in a data frame, choose columns and create a column with the ORF name extracted from the barcode identifiers. Replace the file name with your actual gpr file name. A **setwd**() command indicating

the “working directory” is useful at the beginning of the R session.

```
gprfname <- "252603510022_8_G100526_12vsownmix.gpr"
gprdata <- read.delim(gprfname, skip = 35, header = TRUE, sep =
"\t", stringsAsFactors = F)
gprdata$ORF <- namesextract(gprdata$ID)
```

4. Find the columns that will be used in the normalization process by using either string identity (**==**) or regular expressions:

```
header = names(gprdata)
block_c = which(header == "Block")
col_c = which(header == "Column")
row_c = which(header == "Row")
id_c = which(header == "ID")
name_c = which(header == "Name")
genename_c = which(header == "GeneName")
ORF_c = which(header == "ORF")
median_ratios_c = which(header == "Median.of.Ratios..635.532.")
SNR_c = grep("SNR", header)
F_median_c = grep("^F[0-9]{3}.Median$", header)
B_median_c = grep("^B[0-9]{3}.Median$", header)
```

5. Extract only a few columns of interest and do simple filtering and calculations for M, the log transformed ratio of red and green signal, $\log_2$(Red/Green) and A, the average intensity of the two signals 0.5*$\log_2$(Red*Green). A is a log-average, the log value for the geometric mean of the two signals. The plot of ratios against average intensity can reveal intensity dependent artifacts that need correction.

```
F_median <- gprdata[F_median_c]
B_median <- gprdata[B_median_c]
```

```
FB_median <- F_median - B_median
FB_median[FB_median < 1] <- 1  #adjust missing values to 1
names(FB_median) <- c("signal_R", "signal_G")
to_norm = data.frame(gprdata[, c(ORF_c, block_c, col_c, row_c, id_c,
name_c,genename_c, SNR_c, median_ratios_c)], FB_median)
```

6. Filter and calculate vaues that have low signal:

```
minSNR = 5  #minimum acceptable signal to noise ratio
goodones <- which(to_norm$signal_R > 1 & to_norm$signal_G > 1)
filterSNRidx <- which(to_norm$SNR.635 > minSNR & to_norm$SNR.532
> minSNR)
```

7. Calculate M and A only on the signals that are worth it, others leave as **NA** (not available):

```
goodx <- intersect(goodones, filterSNRidx)
to_norm$M <- NA
to_norm$A <- NA
to_norm$M[goodx] <-
log2(to_norm$signal_R[goodx]/to_norm$signal_G[goodx])
to_norm$A[goodx] <- 0.5 * log2(to_norm$signal_R[goodx] *
to_norm$signal_G[goodx])
```

8. Split the data by group of barcodes (separate for up and down, since coming from two separate PCR reactions).

```
ups <- grep("-U", to_norm$ID)
downs <- grep("-D", to_norm$ID)
to_norm$UD <- NA
to_norm$UD[ups] <- "U"
to_norm$UD[downs] <- "D"
to_norm$UD <- as.factor(to_norm$UD)
```

9. Create a diagnostic plot for the two classes of barcodes (**Fig. 2a**):

```
library(ggplot2)
ggplot(to_norm[goodx, ], aes(x = A, y = M, color = UD)) +
facet_grid(UD ~ .) + theme_bw() + geom_point(alpha = 0.33) +
geom_smooth(method = "loess", size = 0.5, color = "black", alpha
= 0.5) + ylim(-3, 6)
```

10. Find the row numbers for cases when signal higher than background:

```
indexU <- intersect(goodx, which(to_norm$UD == "U"))
indexD <- intersect(goodx, which(to_norm$UD == "D"))
```

11. Simple normalization, by median substraction, will not work well in this case, since the dependence between signal intensity and ratio changes is not linear. Lowess, or loess is a more versatile method. First, define the function, based on the internal "**loess**" function:

```
loessnorm <- function(M, A) {
    loessfit <- loess(M ~ A, span = 0.8) #span is adjustable
    normM <- M - predict(loessfit)
    return(normM)
}
```

12. Apply **loess** normalization to the two classes of signal (U/D):

```
normU <- loessnorm(to_norm$M[indexU], to_norm$A[indexU])
normD <- loessnorm(to_norm$M[indexD], to_norm$A[indexD])
to_norm$normML <- NA  #create a new column
to_norm$normML[indexU] <- normU
to_norm$normML[indexD] <- normD
```

13. Verify that the normalization changed the distribution of the values (**Fig. 2b**):

```
ggplot(to_norm[goodx, ], aes(x = A, y = normML, color = UD)) +
```

```
facet_grid(UD ~ .) + theme_bw() + geom_point(alpha = 0.33) +
geom_smooth(method = "loess", size = 0.5, color = "black") +
ggtitle("Normalized") + ylim(-3, 6)
```

14. Recalculate the green and red signal values from the normalized ratios (if needed):

```
to_norm$norm_R <- sqrt(2^(2 * to_norm$A + to_norm$normML))
to_norm$norm_G <- sqrt(2^(2 * to_norm$A - to_norm$normML))
```

15. Combine the results for each UP and DOWN values, for each ORF and, next, combine the obtained UP and DOWN medians for each ORF.

```
aggregated_UD <- aggregate.data.frame(to_norm[, c("normML",
"A")], by = list(ORF = to_norm$ORF, UD = to_norm$UD), median,
na.rm = T)
aggregated_ORF <- aggregate.data.frame(aggregated_UD[, c("normML",
"A")], by = list(ORF = aggregated_UD$ORF),median, na.rm = T)
```

16. Read a list of genomic positions for the different features of SGD (simple text file, tab delimited) and add this information to the normalized table of values:

```
genopos <- read.delim("ordre_geno.txt", stringsAsFactors = F)
aggregated_ORFgp <- merge.data.frame(aggregated_ORF, genopos,
by.x = "ORF",
by.y = "orf")
```

17. Estimate the presence of the “exclusion region” in the position that corresponds to the gene of interest, in this case the deleted gene is the 4057th feature in the list (**Fig. 3**, *see* **Note 13**):

```
ggplot(aggregated_ORFgp, aes(x = ordre_geno, y = -normML)) +
theme_bw() + geom_point(alpha = 0.5) + geom_vline(xintercept = 4057,
alpha = 0.5, col = "blue")
```

18. The various computations can be better organized in a series of functions to be sequentially applied to the data. Interactive R sessions are best done with RStudio (www.rstudio.com/). For further data processing or analysis, the results can be exported to a comma separated text file:

```
write.csv(aggregated_ORFgp, "aggregated_ORFgp.csv", row.names =
F)
```

3.6 Visualization of series of GIM results

1. To be able to visualize the data from several GIM screens at the same time, build a tabulated file that has a format accepted by Cluster 3 (http://bonsai.hgc.jp/~mdehoon/software/cluster/). It includes a first column with unique identifiers for the genes or mutations followed by an optional, column with annotations (gene name, SGD description line) labelled “NAME” and columns corresponding to normalized numerical values (usually $\log_2$ transformed) for different GIM screens. Cluster 3 has a graphical user interface that allows to keep only rows for which values were obtained in at least 70% of the experiments. Proceed with hierarchical clustering and test several methods. We obtained good functionally relevant grouping of data with 'Correlation (uncentered)' as the similarity metric and with 'Average linkage' as clustering technique.

2. Open the clustering result (.cdt file) with Java TreeView (http://jtreeview.sourceforge.net/) and use either the tabular color-coded view of the cluster or scatterplots (Analysis>Scatterplot) to identify interesting correlations between genetic interaction profiles or similarity in response of the tested mutants as measured in several different GIM screens.

3. More involved data analysis depends on the integration of data from several

different sources (SGA, GIM, protein-protein interactions) and gene ontology (GO) enrichment analysis in groups of related genes (by using, for example, the generic GO term finder, http://go.princeton.edu/cgi-bin/GOTermFinder).

4 Notes

1. Agilent microarrays for barcodes, unlike those used for transcriptome studies, can be stripped and reused, since hybridization is done on a relatively short, 20 nt, region (instead of 60 nt, as usual for expression microarrays). If slides have been previously treated with the Agilent stabilization and drying solution, they should be kept for 4-5 hours in a Tris HCl 10mM, EDTA 1 mM, pH 8 solution at 45°C to remove the protective coating. Stripping is performed by incubating the slides twice for 5 minutes in a solution of 1% SDS, 5 mM Tris HCl pH 8, 0.5 mM EDTA that was brought to 95°C. The stripped slides should be re-scanned to verify that no signal is present after the procedure.

2. The collection of yeast mutants has been distributed to many laboratories in the yeast community under the form of 96 well plates. Single strains should be checked thoroughly for the absence of the deleted gene and for the presence of the KanMX4 cassette at the appropriate locus. Cross-contamination is not a problem when working with pools of mutants, where we can follow a deleted locus via the molecular barcodes, with the exception of slow growing mutants that could be “invaded” readily by faster growing contaminating strains.

3. Thawing plates leads to viability loss; however since there is a large excess of viable cells in a typical well of the collection plates, re-freezing plates allows an economy of plates and manipulation time to be made.

4. Many viable yeast deletion strains have been previously shown to accumulate secondary mutations that alleviate the growth defect of the original strain ***(17)***. The extent of this phenomenon is remarkable, as it was estimated that more than half of the strains from the collection are heterogeneous mixtures of cells bearing different genome variants as a result of the original gene deletion ***(18)***.

5. The presence of two markers on the plasmid used for diploid selection eases selection of transformed yeast cells on SC-URA plates, while culture and selection of diploids is done in rich medium with hygromycine.

6. For a single screen we use an estimated number of $3*10^8$ mutant cells. Since only a fraction of the cells stored at -80°C is viable and the mating efficiency is variable, we estimate the number of diploids obtained to be around $5*10^7$, which corresponds to around 10,000 heterozygous double mutants per locus, assuming an uniform distribution for the number of cells for each mutant in the pool of cells. These relatively large numbers tend to minimize between-screen variation. Another step in which the number of cells is critical is sporulation, since this step is particularly inefficient in cells derived from S288c strain ***(19)***.

7. If some of the very rich GNA medium gets carried in the sporulation medium, an important decrease in the sporulation efficiency can be observed. Thorough wash of diploid cells is thus mandatory for the efficiency of this step.

8. Growth rate estimates are useful to ensure that the query screen pool of double mutants and the reference pool of double mutants grow for a comparable number of generations. If, for example, the doubling time of a

pool of mutants is 2.5 hours, cells should be recovered after 45 hours from the time the antibiotics have been added to the culture. Such timing can be difficult to follow, and an automatic system for cell culture can be set up in such a way that several screens are performed in parallel. The system involves the use of a turbidostat, for which readings of absorbance (proportional with the turbidity and the total number of cells) are done periodically, and defined amounts of fresh media are automatically added to ensure continuous dilution of the culture.

9. We use two separate PCR reactions for barcode amplification and labeling to avoid the massive expansion of the barcode regions from the query strain deletion that is present in all the cells of the population. The KU and KD oligonucleotides are complementary to regions of the KanMX4 cassette that have no equivalent in the MFα2-NAT cassette. While we used to gel purify the products of the first PCR reaction, a simple dilution before the second PCR reaction is sufficient to yield excellent hybridization results. It is important, especially for the second PCR reaction, to set up control PCR reactions, without template DNA. Contamination with PCR products from previous screens is a common occurrence.

10. In general we use the deletion of YEL068C as a reference strain. We observed no specific effects of this deletion on cell growth in previous GIM screens or under various stress conditions ***(20)***. We have also used successfully mixes of genomic DNA obtained from 16 different GIM screens as a base for comparison with a given gene deletion screen.

11. Oligonucleotides used at this step in great excess hybridize with the universal sequences found on both sides of the barcode region, so that only

that region remains available for hybridization with the complementary regions on the barcode microarray.

12. In the absence of a specific rotating system for Agilent hybridization chambers, we adapted a rotating wheel by carving slots of the right size in a poly-styrene foam pad. The hybridization chambers are placed and secured with elastic bands and the rotating wheel is left in a 24°C incubator overnight.

13. The observed peak, centered on the position of the query strain deletion, correspond to gene deletions that are physically close on the chromosome. If the two markers are on the same chromosome, to find both of them in a single cell, a meiotic recombination event must occur. It is possible to use simple probabilistic rules and the Haldane transformation to correct the observed values and estimate genetic linkage distances, as explained in the supplementary information of the initial GIM paper ***(8)***. Briefly, we can estimate the recombination frequency $r = 1/(1+query/reference)$, where the ratio query/reference is the one obtained in the GIM screen after normalization. Haldane genetic distance is found as $ln(1-2r)/2$ and shows a linear correlation with the known physical distance between the markers (Fig. **S1E** in the mentioned supplementary information). Once a fit has been made, we correct the observed values for the effect of genetic linkage. The same method has been recently used to estimate changes in crossing over rates at different chromosomal positions, based on the results of large numbers of SGA screens ***(21)***. Peaks can be also inspected visually, for single GIM screens and the data having obvious biases can be removed from the final table.

Acknowledgments

The protocol presented here was set up initially by Laurence Decourty, Gwenael Badis-Breard and Alain Jacquier, data management and analysis were performed by Christophe Malabat, and large scale developments were made possible by the ANR funding of the GENO-GIM project (ANR-2008-JCJC-0019-01).

References

1. G. Giaever, A.M. Chu, L. Ni, et al. (2002) Functional profiling of the Saccharomyces cerevisiae genome, Nature. 418, 387–391.
2. E.A. Winzeler, D.D. Shoemaker, A. Astromoff, et al. (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis, Science (New York, N.Y.). 285, 901–906.
3. S.L. Ooi, D.D. Shoemaker, and J.D. Boeke (2003) DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray, Nature Genetics. 35, 277–286.
4. X. Pan, D.S. Yuan, D. Xiang, et al. (2004) A robust toolkit for functional profiling of the yeast genome, Molecular Cell. 16, 487–496.
5. C.C. Lindegren and G. Lindegren (1943) A New Method for Hybridizing Yeast, Proceedings of the National Academy of Sciences of the United States of America. 29, 306–308.
6. A.H. Tong, M. Evangelista, A.B. Parsons, et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants, Science (New York, N.Y.). 294, 2364–2368.
7. A.H.Y. Tong, G. Lesage, G.D. Bader, et al. (2004) Global mapping of the yeast genetic interaction network, Science (New York, N.Y.). 303, 808–813.
8. L. Decourty, C. Saveanu, K. Zemam, et al. (2008) Linking functionally related genes by sensitive and quantitative characterization of genetic interaction profiles, Proceedings of the National Academy of Sciences of the United States of America. 105, 5821–5826.
9. M. Schuldiner, S.R. Collins, N.J. Thompson, et al. (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile, Cell. 123, 507–519.

10. M. Costanzo, A. Baryshnikova, J. Bellay, et al. (2010) The genetic landscape of a cell, Science (New York, N.Y.). 327, 425–431.
11. D. Muhlrad and R. Parker (1999) Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance, RNA (New York, N.Y.). 5, 1299–1307.
12. L. Decourty, A. Doyen, C. Malabat, et al. (2014) Long open reading frame transcripts escape nonsense-mediated mRNA decay in yeast, Cell reports. 6, 593–598.
13. F. Battke, S. Symons, and K. Nieselt (2010) Mayday - integrative analytics for expression data, BMC Bioinformatics. 11, 121.
14. A. Floratos, K. Smith, Z. Ji, et al. (2010) geWorkbench: an open source platform for integrative genomics, Bioinformatics (Oxford, England). 26, 1779–1780.
15. J. Cox and M. Mann (2012) 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data, BMC Bioinformatics. 13, S12.
16. J. Hill, K.A. Donald, D.E. Griffiths, et al. (1991) DMSO-enhanced whole cell yeast transformation, Nucleic acids research. 19, 5791.
17. T.R. Hughes, M.J. Marton, A.R. Jones, et al. (2000) Functional discovery via a compendium of expression profiles, Cell. 102, 109–126.
18. X. Teng, M. Dayhoff-Brannigan, W.-C. Cheng, et al. (2013) Genome-wide consequences of deleting any single gene, Molecular cell. 52, 485–494.
19. G. Ben-Ari, D. Zenvirth, A. Sherman, et al. (2006) Four linked genes participate in controlling sporulation efficiency in budding yeast, PLoS genetics. 2, e195.
20. M.E. Hillenmeyer, E. Fung, J. Wildenhain, et al. (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes, Science (New York, N.Y.). 320, 362–365.

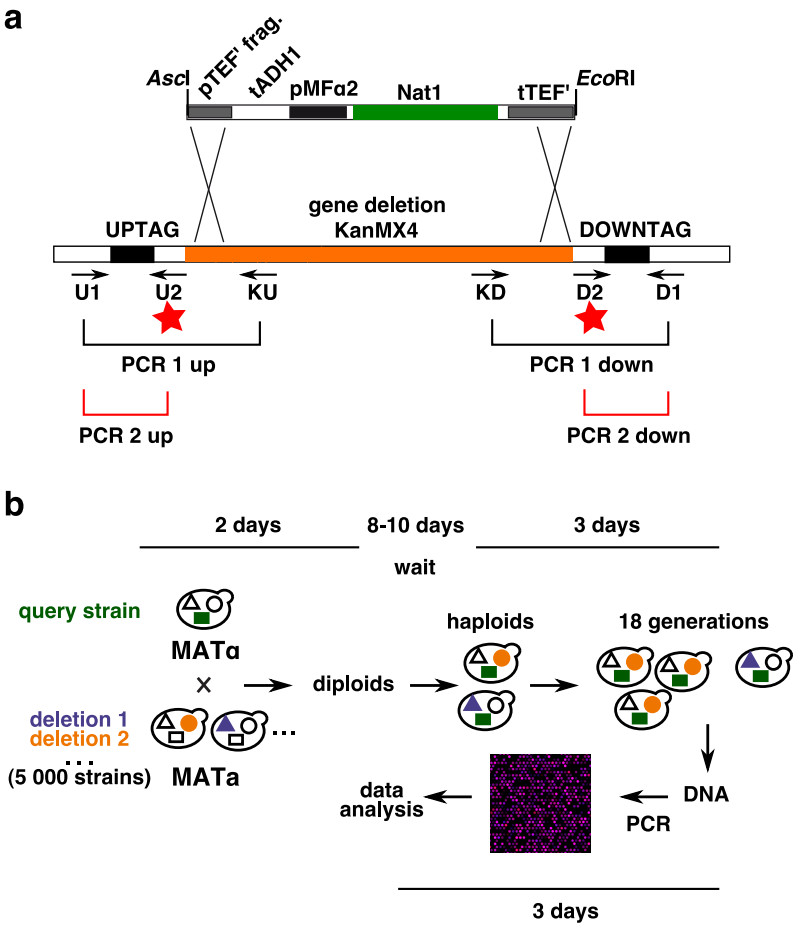
21. A. Baryshnikova, B. VanderSluis, M. Costanzo, et al. (2013) Global linkage map connects meiotic centromere function to chromosome size in budding yeast, G3 (Bethesda, Md.). 3, 1741–1751.

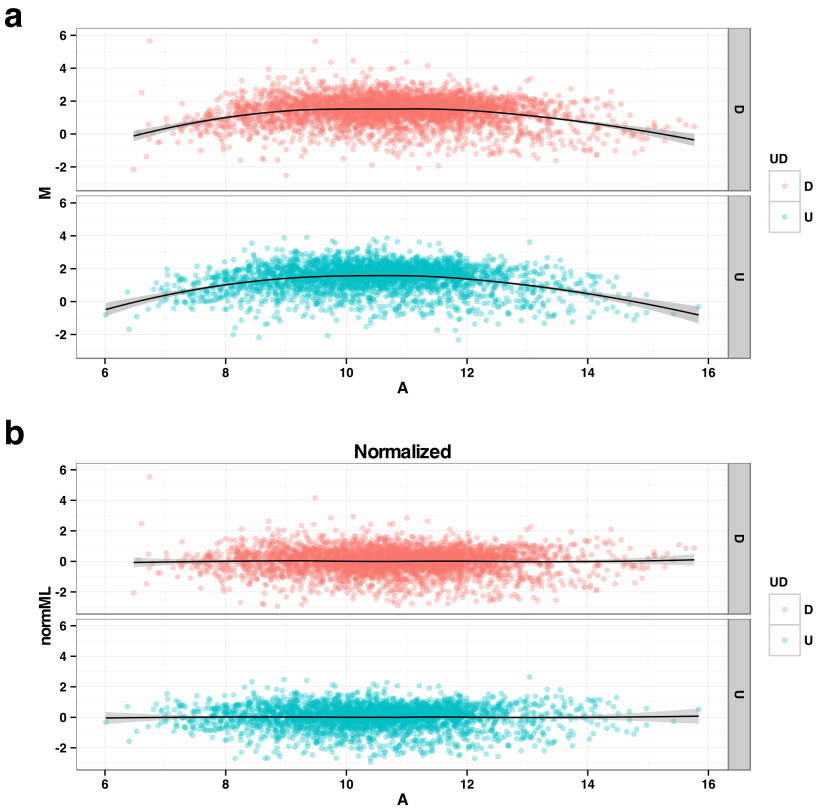
Figure legends

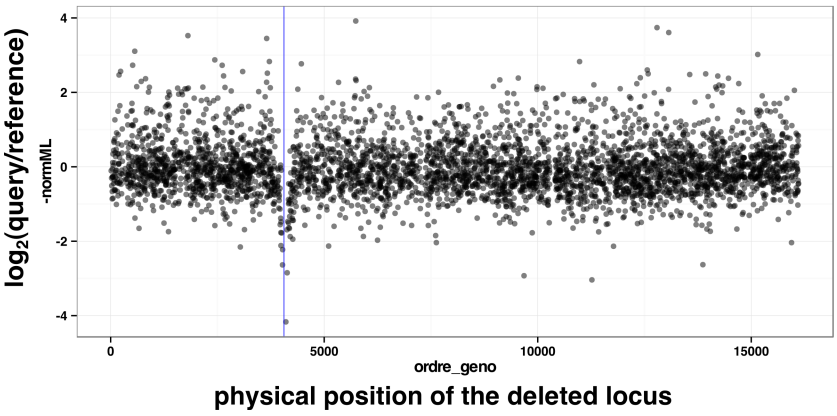
Figure 1. a) The haploid-specific nourseothricin resistance cassette (generated by digestion of pGID3 with *Asc*I and *Eco*RI) replaces the KanMX4 cassette from a *MAT*α strain by homologous recombination with fragments of pTEF' and tTEF'. The barcodes from the KanMX4 strains are amplified in two steps. The second step includes fluorescent labeled oligonucleotides (indicated by red star). **b)** Timing for a typical GIM screen includes 2 days for the obtention of diploid strains and about one week for the selection, processing and microarray analysis of haploid double mutants. Eight to ten additional days are required for the sporulation step.

Figure 2. a) Diagnostic MA plot to estimate variations in the observed signal ratios (M, $\log_2$ transformed) with the intensity of the signal (A, average of signals for the two fluorescence channels, $\log_2$ transformed). **b)** MA plot after **loess** normalization.

Figure 3. Diagnostic plot showing measured double deletion strains growth as a function of the physical position of the tested loci along the yeast chromosomes. The characteristic exclusion peak is marked with a blue vertical line and corresponds to the position of the query gene locus.


a
AscI
pTEF' frag.
tADH1
pMFα2
Nat1
tTEF'
EcoRI
gene deletion
KanMX4
UPTAG
DOWNTAG
U1
U2
KU
KD
D2
D1
PCR 1 up
PCR 1 down
PCR 2 up
PCR 2 down
b
2 days
8-10 days
3 days
wait
query strain
MATα
×
deletion 1
deletion 2
...
(5 000 strains)
MATa
diploids
haploids
18 generations
DNA
PCR
data
analysis
3 days


a
M
A
D
U
UD
D
U
b
Normalized
normML
A
D
U
UD
D
U


log2(query/reference)
-normMIL
4
2
0
-2
-4
0
5000
10000
15000
ordre_geno
physical position of the deleted locus