Electronic Supplementary Material (ESI) for Molecular BioSystems.

Confirmation of the cellular targets of benomyl and rapamycin using next-generation sequencing of resistant mutants in *S. Cerevisiae*

Dustin A. Wride, Nader Pourmand, Walter M. Bray, Tiffany K. Quan Jacob J. Kosarchuk, Sean C. Nisam, Ray F. Berkeley, Sol Katzman, Grant A. Hartzog, Calos E. Dobkin, R. Scott Lokey

slokey@ucsc.edu

Supplementary information

Table of content

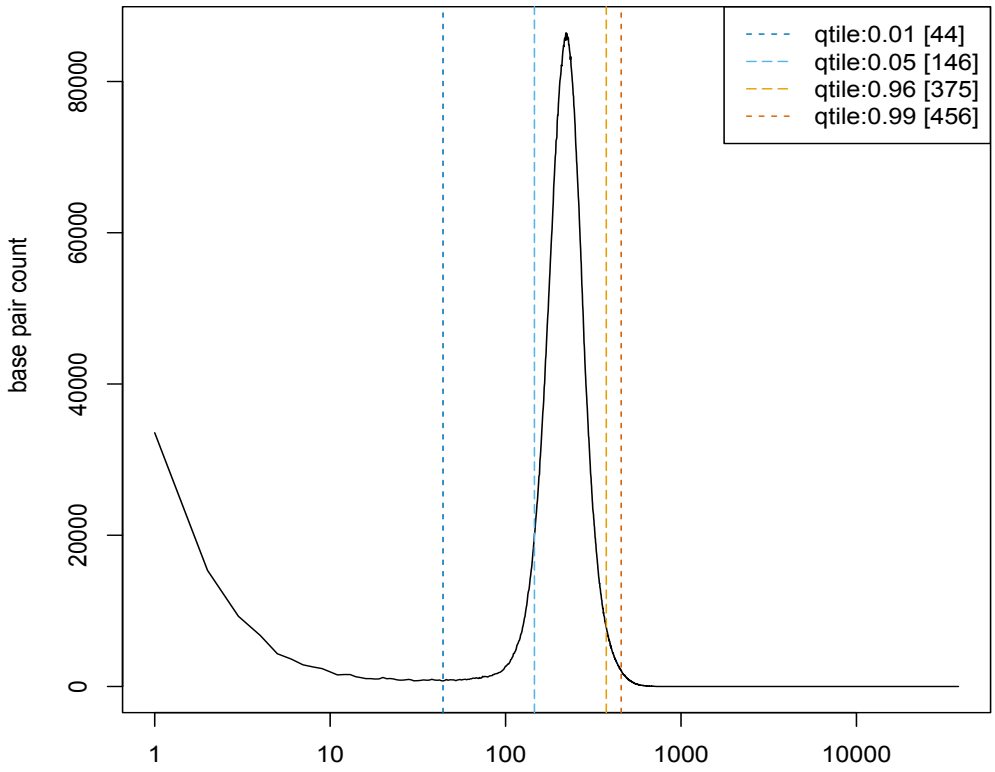
Yeast strain genotype: BY4741 *MAT**a** his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 pdr1Δ0::kanMX4*

Table S1

Compound	Concentration
Hexachlorophene	1mM
Bifonazole	0.5mM
Benzalkonium chloride	2mM
Cetylpyridinium chloride	2mM
Procloroperazine edisylate	100uM
Tetrachloroisophthalonitrile	0.5mM
Chlorhexidine	10mM
Cycloheximide	1mM
Trifluoroperazine HCl	10mM
Chloroxine	10mM
Dihydocelastrol	10mM
Tioconazole	100uM
Cetrimonium Bromide	5mM
Clotrimazole	35mM

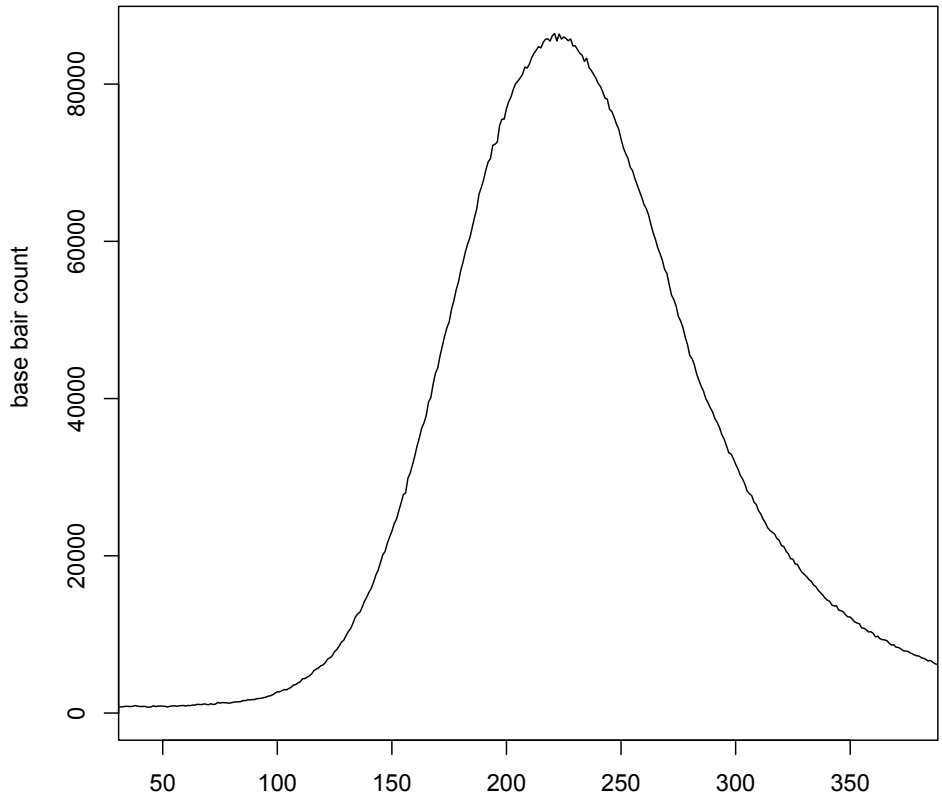
Figure S1

A



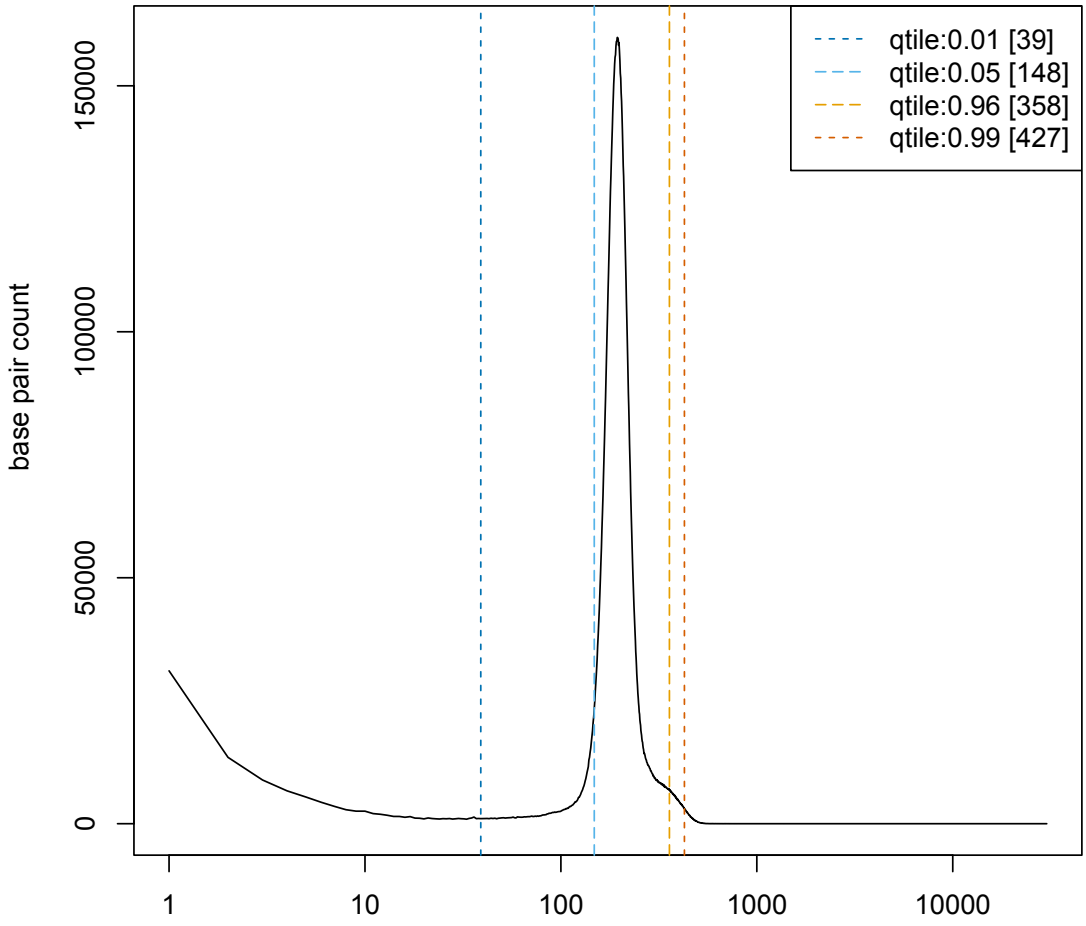
Genome-wide coverage distribution for pdr1 knockout parental control. Log scale read depth values (x axis) for all bases with at least one read (95% of genome). These were filtered further to include only those within the 1% (44 reads) and 96% (375 reads) boundries.

B



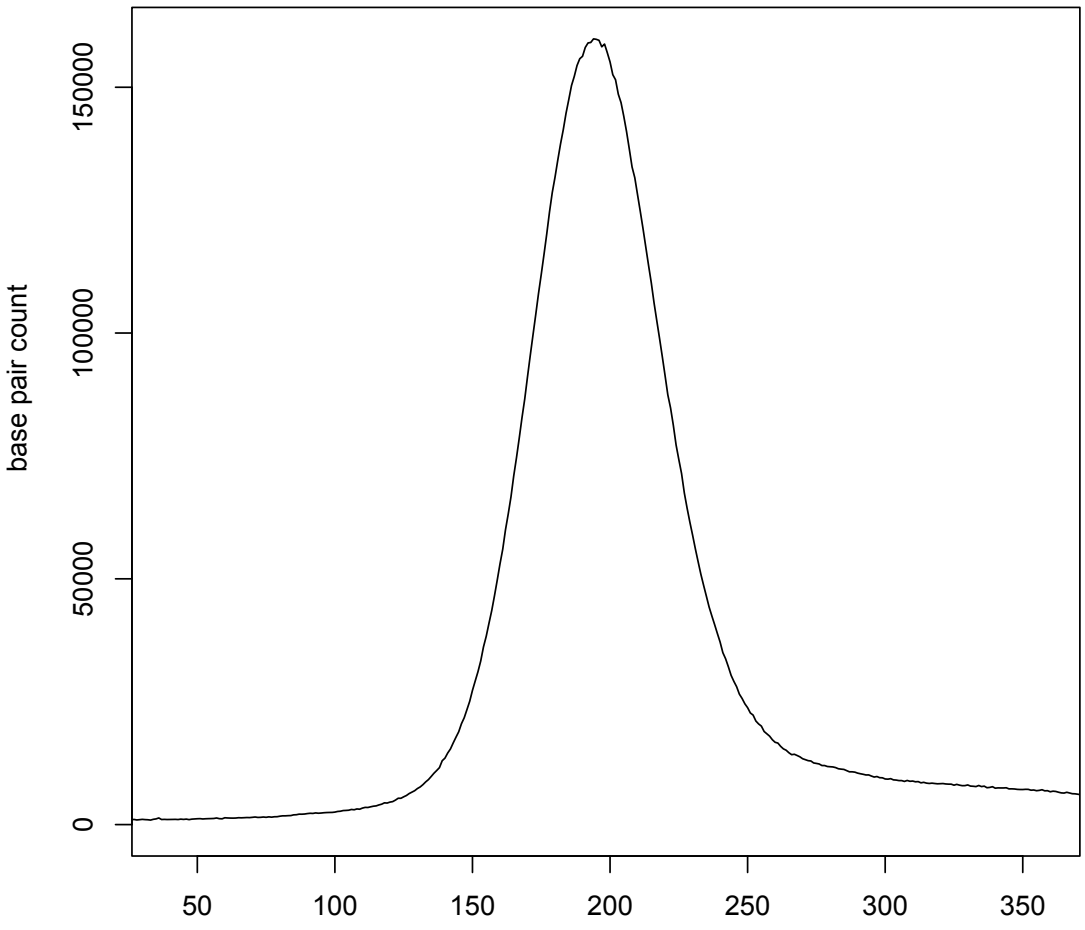
Pdr1 trimmed coverage distribution. Linear scale read depth values (x axis) for all bases between the 1% and 96% boundries for the parental control. Median=229, mean=232, standard deviation=53.7.

C



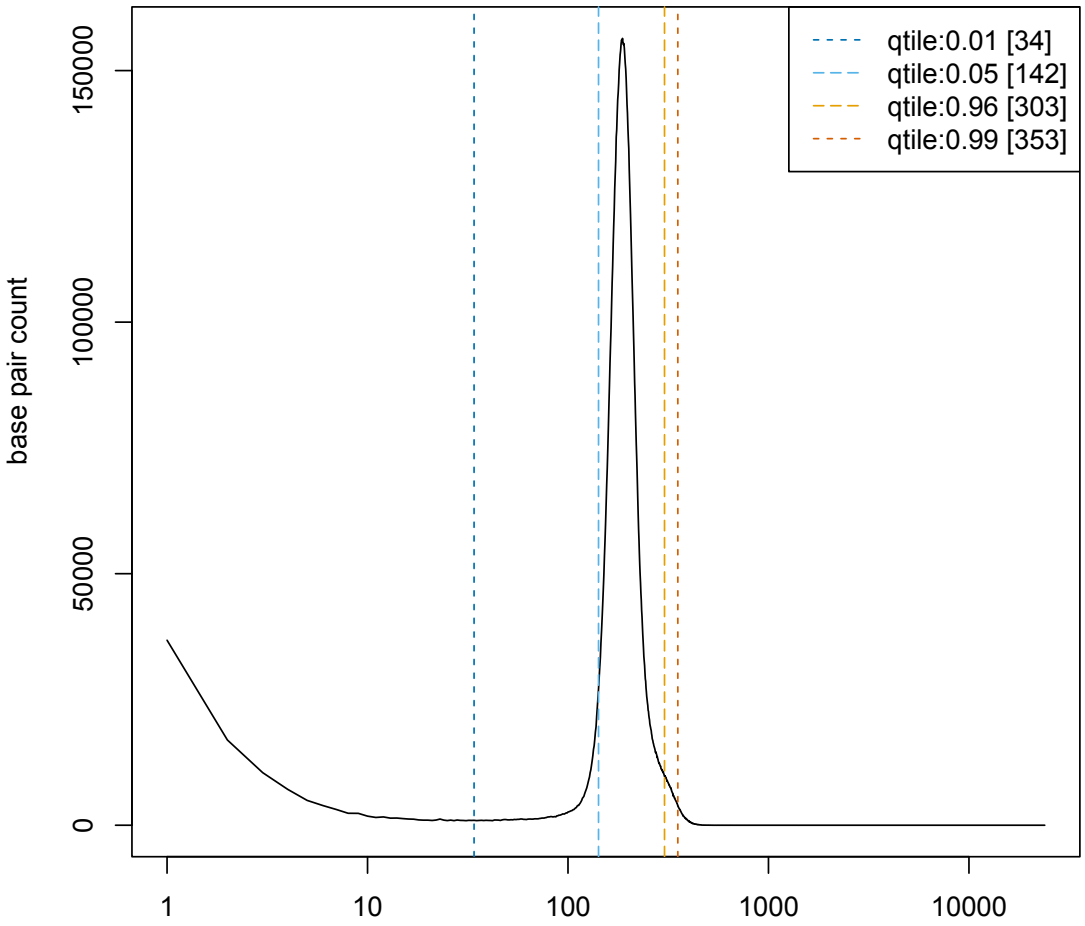
Genome-wide coverage distribution for the benomyl pool. Log scale read depth (x axis) values for all bases with at least one read (95% of genome). These were filtered further to include only those within the 1% (39 reads) and 96% (358 reads) boundries.

D



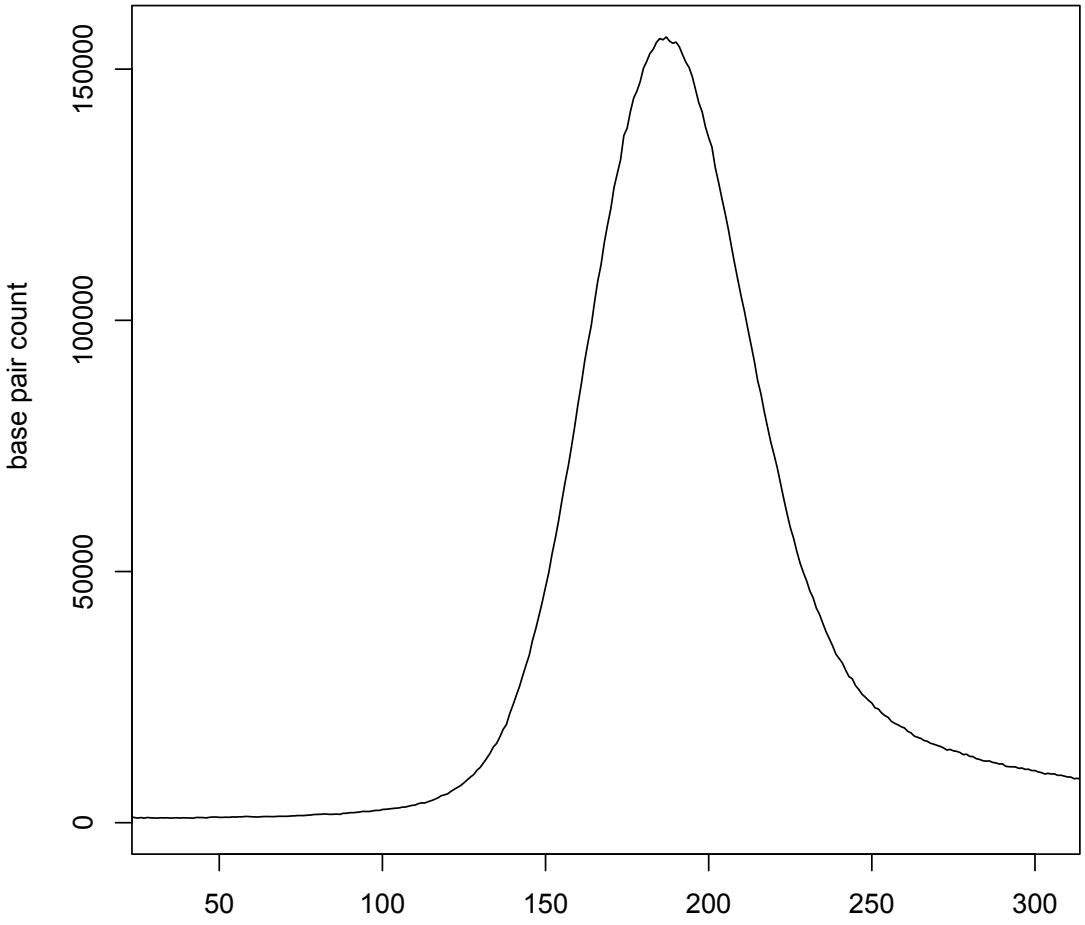
Benomyl trimmed coverage distribution. Linear scale read depth values (x axis) for all bases between the 1% and 96% boundries for the benomyl pool. Median=198, mean=204 and standard deviation=41.9.

E



Genome-wide coverage distribution for the rapamycin pool. Log scale read depth (x axis) values for all bases with at least one read (95% of genome). These were filtered further to include only those within the 1% (34 reads) and 96% (303) boundries.

F



Rapamycin trimmed coverage distribution. Linear scale read depth values (x axis) for all bases between the 1% and 96% boundries for the rapamycin pool. Median=191, mean=194 and standard deviation=35.4.

Probability that x or more strains in a pool have SNPs at a single base

In this section we describe how we estimate the probability that x or more strains have SNPs at a single base somewhere in the genome in the absence of selection. We do this under the assumption that the mutations we observe occur with equal probability at every base. Finding more strains with mutations at a particular base than are likely to occur by chance suggests that the mutation is being selected for. Let S to be the total number of strains included in the pool (in the benomyl pool S=9 and in the rapamycin pool S=5). Define X_A to be the total number of strains with a mutation at a particular base where the reference strain is an A and p_A the probability of a mutation at a base A. The probability of getting X_A strains with a mutation at a single base is

$$P(X_A) = (1 - p_A)^{(S - X_A)} p_A^{X_A} \binom{S}{X_A}.$$

The probability of having x or more strains with a SNP at particular location where the reference strain has an A base is $\sum_{j=x}^{S} P(j)$. As a result the probability that the maximum value that X_A takes at any base in the genome ($X_{A,max}$) is greater than or equal to x is

$$P(X_{A,max} \geq x) = 1 - \left(1 - \sum_{j=x}^{S} P(j)\right)^{b}$$

where b is the number of exonic A bases in the reference genome. A similar formula will generate the estimates for the other three bases. To determine the probability of getting x or more mutations at a base somewhere in the genome we estimate the probability of the complement of that event (getting less than x for all four bases) then subtract it from one.

$$P(X_{max} \geq x) = 1 - P(X_{A,max} < x) * P(X_{G,max} < x) * P(X_{C,max} < x) * P(X_{T,max} < x) \qquad (1)$$

For both the benomyl and rapamycin the probability of getting three or more strains with a SNP at a single base anywhere in the genome is very low as can be seen from the estimates of equation 1 in the table below. The estimates differ substantially between the two pools due to the fact that the Rapamycin pool is smaller and has a much lower mutation rate. The estimates are not very sensitive to reducing the number of bases on which a mutation can fall to simulate hot spots.

Probability x or more strains in a pool have SNPs at a single base

x	Benomyl (N=9)	Rapamycin (N=5)
2	0.17	0.00044
3	1.8E-05	1.4E-09
4	1.2E-09	0

Probability of a GL-AC of X or more

In this section we determine how likely we are to get GL-ACs[1] as large or larger than the ones we observe if the SNPs we observe in the experiment are distributed at random across genes, which is what we would expect in the absence of selection. To conduct the analysis we created a file with the base pair composition of each gene. We then distributed the SNPs we observed in the experiment across the genes with the sampling probabilities determined by the length and base composition of the gene. The GL-AC was computed by summing the allele count of all the SNPs that fell on each gene. We repeated this 100,000 times then determined how frequently the maximum GL-AC was greater than x which is what we present in the table below.

Probability Maximum GL-AC for any Gene is $\geq x$

x	Benomyl (N=9)	Rapamycin (N=5)
0	1	1
1	1	1
2	1	0.271
3	1	0.002
4	1	0
5	0.650	0
6	0.187	0
7	0.040	0
8	0.008	0
9	0.002	0

Note: These are results from 100,000 draws with the same distribution of allele counts at each base as observed in the experiment. p-values differ between drugs/targets due to differences in pool size, mutation rate and gene length and composition

The table reveals that the GL-AC of 8 we found for TUB2 is an unlikely event as a GL-AC of 8 or greater occurred on in only about 8 of every thousand simulations. The two genes with the next highest GL-ACs in the benomyl pool had GL-ACs of 5 which we would expect to occur in 65 percent of experiments in the absence of selection. The code to implement either analysis in R is available on request.

[1]The GL-AC for gene g is GL-AC$_g = \sum_{\text{base i}\in g} AC_i$.